GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 12:59:04 ; Search time 5686.68 Seconds
(without alignments)
5067.252 Million cell updates/sec

Title: US-09-997-610-3
Perfect score: 1377
Sequence: 1 athgtngtnathcngtnyt.........gyytntayytncaytcgyatg 1377

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   GenEmbl:*
```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	921	66.9	145880	9	HS302D9	Z82198 Human DNA s
2	497.8	36.2	154090	9	AC025577	AC025577 Homo sapi
3	497.8	36.2	186660	9	AC026107	AC026107 Homo sapi
4	493.6	35.8	124518	9	AL138962	AL138962 Human DNA
5	493	35.8	123070	9	AC008799	AC008799 Homo sapi
6	493	35.8	189768	9	AC044889	AC044889 Homo sapi
7	490.4	35.5	198545	9	AC017063	AC017063 Homo sapi
8	488.8	35.5	161575	9	AC025233	AC025233 Homo sapi
c 9	488.8	35.5	163674	9	AC087500	AC087500 Homo sapi
10	488.8	35.5	175639	9	AC015727	AC015727 Homo sapi
11	488.6	35.5	181842	9	AL391823	AL391823 Homo sapi
12	488.2	35.5	77405	2	AL353634	AL353634 Homo sapi
13	488.2	35.4	124531	9	HSJ633H17	AL049710 Human DNA
14	487.6	35.4	40714	9	Z69722	Z69722 Human DNA s
15	487.6	35.4	159712	9	AP004219	AP004219 Homo sapi
c 16	487	35.4	122295	2	AL672061	AL672061 Homo sapi
17	487	35.4	146743	9	AC093588	AC093588 Homo sapi
18	487	35.4	152544	9	CNS0STEJ	CNS0STEJ Human chr
19	487	35.4	166679	9	AL359232	AL359232 Human chr
20	487	35.4	174662	2	AC026036	AC026036 Homo sapi
21	487	35.4	176426	9	AC007370	AC007370 Homo sapi
c 22	487	35.4	203726	2	AC011882	AC011882 Homo sapi
c 23	486	35.3	139005	2	AF286112	AF286112 Homo sapi
c 24	486	35.3	156458	2	AC025647	AC025647 Homo sapi
25	486	35.3	165901	9	AP003474	AP003474 Homo sapi
c 26	486	35.3	171427	9	AC021590	AC021590 Homo sapi
c 27	485.8	35.3	99084	2	AC026420	AC026420 Homo sapi
c 28	485.8	35.3	137790	9	AC010280	AC010280 Homo sapi
c 29	485.8	35.3	150332	9	AC004921	AC004921 Homo sapi
c 30	485.8	35.3	174293	9	AC008816	AC008816 Homo sapi
31	485.8	35.3	203043	9	AC017093	AC017093 Homo sapi
32	485.4	35.3	131215	9	AC079614	AC079614 Homo sapi
33	485	35.2	107885	9	AC006389	AC006389 Homo sapi
c 34	484.6	35.2	174480	9	CNS00M8T	CNS00M8T Human chr
35	483	35.1	164370	9	AC097460	AC097460 Homo sapi
c 36	482.6	35.0	112515	9	AL139216	AL139216 Human DNA
37	482.6	35.0	142224	2	AC096535	AC096535 Homo sapi
38	482	35.0	178555	9	AL354978	AL354978 Homo sapi
c 39	482	35.0	168502	9	AC091005	AC091005 Homo sapi
c 40	481.6	35.0	38235	9	AC004559	AC004559 Homo sapi
c 41	481.4	35.0	234622	2	AC110089	AC110089 Homo sapi
c 42	479.8	34.8	187883	9	AC008268	AC008268 Homo sapi
c 43	479.8	34.8	167891	9	AC013439	AC013439 Homo sapi
c 44	479.2	34.8	73390	9	AC012038	AC012038 Homo sapi
c 45	479.2	34.8	94730	9	AP000230	AP000230 Homo sapi

ALIGNMENTS

```
RESULT   1
HS302D9
LOCUS       HS302D9               145880 bp    DNA     linear   PRI 12-DEC-1999
DEFINITION  Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
            GSSs, complete sequence.
ACCESSION   Z82198
VERSION     Z82198.2  GI:6572207
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145880)
AUTHORS     Bridgeman,A.
TITLE       Direct Submission
JOURNAL     Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 13, 1999 this sequence version replaced gi:3164067.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
```

corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RP1-302D9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-302D9 The true left
end of clone CTA-282F2 is at 69682 in this sequence. The true right
end of clone CTA-415C2 is at 55167 in this sequence.

```
FEATURES             Location/Qualifiers
source          1..145680
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22"
                /clone="RP1-302D9"
                /clone_lib="RPCI-1"
repeat_region   188..245
                /note="MER3 repeat: matches 144..209 of consensus"
repeat_region   246..571
                /note="AluSx repeat: matches 1..312 of consensus"
repeat_region   572..759
                /note="MER3 repeat: matches 1..144 of consensus"
repeat_region   783..933
                /note="MER5A repeat: matches 26..187 of consensus"
repeat_region   1033..1336
                /note="AluSp repeat: matches 1..299 of consensus"
repeat_region   1450..1583
                /note="MIR repeat: matches 24..160 of consensus"
repeat_region   1687..1752
                /note="L2 repeat: matches 2593..2661 of consensus"
repeat_region   2350..2660
                /note="AluSc repeat: matches 3..309 of consensus"
repeat_region   2684..2981
                /note="AluSq repeat: matches 2..300 of consensus"
repeat_region   3323..3343
                /note="MIR repeat: matches 116..136 of consensus"
repeat_region   3344..3652
                /note="MLT1E repeat: matches 359..568 of consensus"
repeat_region   3653..3928
                /note="AluY repeat: matches 1..309 of consensus"
repeat_region   3929..4278
                /note="MLT1E repeat: matches 136..359 of consensus"
repeat_region   4279..4485
                /note="THE1B repeat: matches 3..364 of consensus"
misc_feature    5073..5176
                /note="52 copies 2 mer ct 78 conserved"
repeat_region   5181..5491
                /note="AluJb repeat: matches 1..311 of consensus"
repeat_region   6369..6485
                /note="L2 repeat: matches 2579..2705 of consensus"
repeat_region   6647..6685
                /note="MADE1 repeat: matches 1..23 of consensus"
repeat_region   6686..6987
                /note="AluSx repeat: matches 1..302 of consensus"
repeat_region   6988..7036
                /note="MADE1 repeat: matches 23..77 of consensus"
repeat_region   7482..7754
                /note="AluUb repeat: matches 9..290 of consensus"
repeat_region   7775..8060
                /note="AluJo repeat: matches 1..295 of consensus"
repeat_region   8414..8551
                /note="L2 repeat: matches 2553..2706 of consensus"
repeat_region   8914..9030
                /note="MIR repeat: matches 147..262 of consensus"
repeat_region   9110..9280
                /note="MIR repeat: matches 91..262 of consensus"
repeat_region   9283..9412
                /note="MIR repeat: matches 15..144 of consensus"
repeat_region   9521..9679
                /note="FAM repeat: matches 3..161 of consensus"
repeat_region   9820..10225
                /note="MSTB repeat: matches 2..425 of consensus"
misc_feature    complement(10179..10678)
                /note="match: GSS: Em:B56592"
misc_feature    complement(10204..10728)
                /note="match: GSS: Em:AQ701486"
misc_feature    complement(10249..10706)
                /note="match: GSS: Em:AQ225495"
repeat_region   10312..10383
                /note="MIR repeat: matches 79..150 of consensus"
misc_feature    10718..11310
                /note="match: GSS: Em:B14024"
misc_feature    10784..11201
                /note="match: GSS: Em:B43656"
repeat_region   11838..11946
                /note="MIR repeat: matches 20..137 of consensus"
repeat_region   12174..12445
                /note="L2 repeat: matches 1988..2275 of consensus"
repeat_region   12444..12642
                /note="AluSp repeat: matches 1..302 of consensus"
misc_feature    13017..13369
                /note="match: STS: Em:G49301"
repeat_region   13331..13397
                /note="MIR repeat: matches 174..244 of consensus"
repeat_region   13398..13698
                /note="AluSp repeat: matches 1..302 of consensus"
repeat_region   13699..13810
                /note="MIR repeat: matches 76..174 of consensus"
repeat_region   13806..13919
                /note="MIR repeat: matches 189..301 of consensus"
repeat_region   13945..14060
                /note="MIR repeat: matches 77..189 of consensus"
repeat_region   14061..14367
                /note="MIR repeat: matches 24..142 of consensus"
repeat_region   14368..14452
                /note="AluY repeat: matches 1..301 of consensus"
repeat_region   14589..14679
                /note="MIR repeat: matches 141..225 of consensus"
misc_feature    14597..15201
                /note="match: GSS: Em:AQ533482"
misc_feature    14616..15060
                /note="match: GSS: Em:AQ370601"
repeat_region   14868..15040
                /note="MIR repeat: matches 173..262 of consensus"
repeat_region   15071..15188
                /note="L2 repeat: matches 2112..2239 of consensus"
repeat_region   15304..15399
                /note="MLT1B repeat: matches 1..99 of consensus"
repeat_region   15490..15662
                /note="AluSg1 repeat: matches 2..114 of consensus"
repeat_region   15669..15727
                /note="MLT1B repeat: matches 119..178 of consensus"
repeat_region   15728..16027
                /note="AluSc repeat: matches 1..299 of consensus"
repeat_region   16028..16245
                /note="L2 repeat: matches 2..216 of consensus"
repeat_region   16546..16854
                /note="MLT1B repeat: matches 178..390 of consensus"
repeat_region   18296..18323
                /note="AluY repeat: matches 1..300 of consensus"
repeat_region   18324..18392
                /note="MSTA repeat: matches 2..29 of consensus"
```

repeat_region   18893. .18712
  /note="MER66-internal repeat: matches 4919. .4993 of
  consensus"
repeat_region   18713. .19133
  /note="AluJb repeat: matches 1. .311 of consensus"
repeat_region   complement(18872. .19230)
  /note="MER66-internal repeat: matches 4548. .4919 of
  consensus"
misc_feature   19251. .19719
  /note="match: GSS: Em:AQ005063"
misc_feature   19937. .20290
  /note="match: GSS: Em:BI4179"
repeat_region   20317. .20382
  /note="HERVH21 repeat: matches 4657. .5784 of consensus"
repeat_region   20513. .20666
  /note="33 copies 2 mer ta 68 conserved"
repeat_region   20682. .21008
  /note="77 copies 2 mer tt 70 conserved"
repeat_region   21239. .21553
  /note="AluSg1 repeat: matches 1. .306 of consensus"
repeat_region   21882. .22254
  /note="HUERS-P3 repeat: matches 4410. .4713 of consensus"
repeat_region   22302. .22537
  /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   22238. .22850
  /note="MER66-internal repeat: matches 2186. .2417 of
  consensus"
repeat_region   22851. .23801
  /note="AluSp repeat: matches 1. .313 of consensus"
repeat_region   23905. .23989
  /note="MER66-internal repeat: matches 1210. .2186 of
  consensus"
repeat_region
  /note="MER66-internal repeat: matches 3017. .3102 of
  consensus"

Query Match   66.9%;   Score 921;   DB 9;   Length 145880;
Best Local Similarity   58.2%;   Pred. No. 2e-211;
Matches 750;   Conservative 300;   Mismatches 239;   Indels 0;   Gaps 0;

```
Qy      89 tnggncnccnggngcncnggngnytnccncaartayacnggngarathwsngaratgacna  148
              |:::|| || ||:||  |||:|| ||:|| ||||||||||||||||
Db   37254 TGGAAATGTCGTGGTGCCCCAGGTTNACCAATATACAGGAGAAAAATAACTGACAA  37313

Qy     149 artgycntgycnngayathgargmgnwsngcnttyacngtnaarytnwsnggnaarytnc  208
Db   37314 AATCCCCCTGTCCTGATATATAGAAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAAACTTC  37373

Qy     209 cnytnccncnttyaarccnathathttyacnggngtnytntayaayvcncarmngayytna  268
Db   37374 CTCTTCCTTTCAAGCCCATCACATCTTCACAGGGGTCCCTGTACAANTGCCCAGAGGGATTTAA  37433

Qy     269 argargcnatggggngtnttycgntcgymgngtnccngnaaytaytaywsnwsnttygayg  328
Db   37434 AGGAGGCCATGGGAGTCTTTGCTTGCAGGGTGCCTGGGAATTACTACTCCAGCTTTGATG  37493

Qy     329 tngarytncaycaycaytgyaargtnaatgmgnaarcaaraththgtngcnaaya  388
Db   37494 TTGAGCTGCGTCATNCATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAAATTTGGCTAATA  37553

Qy     389 arcgargarathwsnaaarcaarcarwsnathcarargayrgargtnacntgggtnytnaaytgcnt  448
Db   37554 AGGAGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGCGTGCTGTTAAAGGCAT  37613

Qy     449 tywsntyathgmgngargcngarcayaarwsnwsngaraayytncaycnngayaaygtna  508
Db   37614 TCAGTTCATTAAGGAGGCGAGAGCGCATAAGCAGTTCAGAAAAATTTGCACCCTGACAATGTGA  37673

Qy     509 thaaraaaaaraayccnttywsngargnaartyaaytngcngnagarathgyatht  568
Db   37674 TAAAAAAGAAAAAACCATTTCTGAGGGGAAATTCAAGCTGGCTGCAGAATTTGCATAT  37733

Qy     569 gyaaygargaarytnaaytcncaargayaaygngaraayaayaagithwsntggacntgyc  628
Db   37734 GTAATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGGAAAAATATCTCCTGGACATGTC  37793

Qy     629 armgnwsnwsncaarcarwsnathaarwsnytngcntggmgnccnmgnmgnaaartggttyt  688
Db   37794 AGAGGTCTTCACAGCAGTCACAAANTCACTGGGCCTGGAGGCCTAGGAGAAAATGGTTTT  37853

Qy     689 gyyggnacnggnccngnwsnytntgytgygtnacarccnmgngayytngtnccntgyggtnc  748
Db   37854 GTGGGACAGGCCCAGGGTCCCCTGTGGTCGTTGTCCCAGCTCAGAGACTTGGTGTCC  37913

Qy     749 cngtnaaywsngcngtngcnwsngargyngngcnwsnccnaarccccntggcarytncccnwsng  808
Db   37914 CAGTTAATTCACCTGTGCGTTCAGCTGTGCGGTGCAAGCCCCCAAGCCTTGCCAGTG  37973

Qy     809 gngtngarccngtnggngcnaararwsnmgnathgargtntgtgatggarccncncnathmgnt  868
Db   37974 GTGTTGAGCCTGTGGGTGCAAGGGTCAAGAAGTTCAAGAAATTGAGGTGGGAACCTCCAAATCATGAT  38033

Qy     869 tycarararathtayggnaayccntggatgcnmgncararaartaywsnmgnaathgargt  928
Db   38034 TTCAGAAGATATATGGAAACCCCTGGATGCCCAGGAGAAGTTTGCTGTGTGCGTAGGGGGT  38093

Qy     929 snwsnytggmgnacnwsngcnngngtngtncararggnaaytnggayytna  988
Db   38094 CCTCACTGGAGAACCTCTGCAAGGGTACTACAAAAGGGAAAAATGTTGGGTGGTGGGGCCCCCAC  38153

Qy     989 aymgngtnccnwsnsggngccnwsnwsnmgngacntgayaatgtcctgaaaayytgay  1048
Db   38154 ACAGAGTCCCCAGTGGGGCTCCATCNTAGTGAGGACTCTGTGAGAAGAAGTCCACCATCCTCCA  38213

Qy    1049 gnytncaargngmgmwsnacngaywsnytncarcaygtnccngaraarwsnacngaya  1108
Db   38214 GACTCCAGAAGGGTAGAATCCACTGACAGCTTCAGCAATGTGCCTGAAAAATCCACAGACA  38273

Qy    1109 cncaartgycarccncngtnaaargcngcnggnatggararwsngtnccntayaaracngtngtng  1168
Db   38274 CTCAGTGCCAGCCTGTCAAAAGCAGCAGGGATGGAGTCTGTACCTACAAAACCGTAGTGG  38333

Qy    1169 cngarytnacnaaragacngtnggnathtayytntaytncaytgycaygayytngnagygtnmgnc  1228
Db   38334 CAGAGCTGACCAAGAACCGTGTGGGAATCTACCTCTTGCATTGYCAYGACCTGACGTGAGAC  38393

Qy    1229 ayggngtnaaarmgngaycaytytgygggncnytnmgnttygaytgyccnacnggngnttymgna  1288
Db   38394 ATGGAGTCAAAAAGAGATCATTTGGGACCTTTAAGATTTGACTGCCCCACTGGAATTCGGA  38453

Qy    1289 cntayatgggncnngtnccnytntgytttyggncaarttyttyccnttyggnacngcngtnt  1348
Db   38454 CTTATATGGGGCGCCCGTACCCCTTTGTTTTTGGCCAATTTTTTCCATTTGGAACTGCCGTAT  38513

Qy    1349 tyacncaartgyytntayytncaytgyatg  1377
Db   38514 TTACCCAAATGCCTGTTACCTCCATTGTATG  38542
```

RESULT   2
AC025577
LOCUS   AC025577   154090 bp   DNA   linear   PRI 25-AUG-2000
DEFINITION Homo sapiens 12 BAC RP11-13C3 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC025577
VERSION   AC025577.15   GI:9910028
KEYWORDS   HTG.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 154090)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

```
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,H., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,F., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Hohnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Muzny,D., Neal,J., Nelson,D., Newton,J.,
            Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E.,
            Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A.,
            Payton,B., Peery,D., Perez,L., Peters,L., Pickens,R., Primus,E.,
            Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I.,
            Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H.,
            Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T.,
            Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P.,
            Tamerisa,A., Tang,H., Tang,H., Tansey,J., Taylor,C., Taylor,T.,
            Telrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V.,
            Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154090)
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 154090)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 25, 2000 this sequence version replaced gi:9664948.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu.
```

```
            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.

            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.

            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.
```

```
            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum of 2 clones and 2
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities with no ambiguities for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

QUALSTAT-REPORT---------

Contig length:                                          154090
Phrap values in estimate:                               153545
----------- Summary Statistics -----------
Average error rate (BCM-Phrap estimate):                0.00261755
Fraction of Phrap values less than 40 :                 0.0284086
Number of consensus changing edits:                     22
Number of N's in consensus :                            0
```

```
----------- Consensus changing edits -----------
Position    Original+Context           Edited+Context
9347        aacacagacc(n)tttttttt       aacacagacc(t)tttttttt
10135       attgccccct(n)taggaaaga      attgccccct(t)taggaaaga
10353       ctgtcaccag(n)aaaagatcca     ctgtcaccag(t)aaaagatcca
10417       agcaatgcag(n)ggctacagaa     agcaatgcag(t)ggctacagaa
61959       cttactattt(n)tgtttattct     cttactattt(t)tgtttattct
61960       ttactattt(n)gtttattcca      ttactattt(t)gtttattcca
61961       ttactattnt(t)gtttattctaa    ttactattt(g)tttttattcta
61982       acaacacaag(n)taggtttgg      acaacacaag(t)ttaggtttgg
61983       ntaggtttgg(n)ttactcttcc     ttaggtttgg(t)ttactcttcc
61993       tcattgaccc(n)ctggtcattc     tcattgaccc(t)ctggtcattc
62205       actgcaacct(n)tgcctcccag     actgcaacct(c)tgcctcccag
62956       tttacatata(n)cagtacagta     tttacatata(a)cagtacagta
63722       cctagaaaaa(n)gacntttct      cctagaaaaa(a)gacntttct
90212       gaaaaagac(n)ttttcttttt      gaaaaagac(t)ttttcttttt
90216       agcaatgac(n)ttttcttttt      agcaatgac(t)ttttcttttt
90228       tttctttttc(n)atcttttcat     tttctttttc(t)tttttttcat
90236       tanttttttc(n)atcttttagg     tattttttc(t)atcttttagg
90443       acttagatg(n)ctcctcccag      acttagatg(a)ctcctcccag
91405       cacagcttaa(n)caaggtaaga     cacagcttaa(a)caaggtaaga
122290      aaaaaaagaa(n)gtaagagaaa     aaaaaaagaa(a)gtaagagaaa
137090      acagagaaag(n)caaaacactc     acagagaaag(a)caaaacactc
141632      tgaagccact(n)gaaaagtaat     tgaagccact(t)gaaaagtaat
```

```
----------- Distribution of Quality < 40 Bases -----------

       1000|
        900|
        800|
        700|
   #    600|
        500|
        400|
  bases 400|
        300|                    *
        300|                  * * *
        200|                * * * *
        100|            * * * * *
        100|          * * * * * *
          0|  *     * * * * * * *
               5  10  15  20  25  30  35  40
                       Phrap Value Range
```

FEATURES   Location/Qualifiers

Version: 1.01 qxfo.

```
source          1  .154090
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="12"
                   /clone="RP11-13C3"

repeat_region   54  .288
                   /rpt_family="MIR"
repeat_region   1020  .1146
                   /rpt_family="MER3"
repeat_region   complement(1248  .1584)
                   /rpt_family="MER46B"
repeat_region   complement(1624  .1682)
                   /rpt_family="MIR"
repeat_region   2630  .2761
                   /rpt_family="MIR"
repeat_region   4088  .4216
                   /rpt_family="GA-rich"
repeat_region   5232  .5301
                   /rpt_family="L2"
repeat_region   complement(5450  .5749)
                   /rpt_family="MER102"
repeat_region   6571  .6616
                   /rpt_family="AluY"
repeat_region   complement(6678  .7320)
                   /rpt_family="(TG)n"
repeat_region   7419  .7454
                   /rpt_family="L1ME3"
repeat_region   complement(7741  .7807)
                   /rpt_family="AT_rich"
repeat_region   8113  .8267
                   /rpt_family="MER5A"
repeat_region   8385  .8587
                   /rpt_family="MER5A"
repeat_region   8589  .8630
                   /rpt_family="MIR"
repeat_region   8666  .8695
                   /rpt_family="L2"
repeat_region   /rpt_family="(TTA)n"
```

```
Query Match
Best Local Similarity    36.2%;
                         47.7%;   Score 497.8;  DB 9;  Length 154090;
Matches  488;  Conservative  216;  Mismatches  282;  Indels  36;  Gaps  2;
```

```
Qy    390 rgargarathwsnaarcarwsnathcarargrgtnacntgggtnytnaargtnaargcntt 449
          :|:||||:|||::|||:||:|| |:|::|||: ||||:||||| |||||||:|:|:|:|:||
Db 101633 GGAGAGAATTTCTAAGTAGCAAAAGCATTAAAGAGGTGACTTAGAATGTGGTTAAAGGTATT 101692

Qy    450 ywsnttyathmgnargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ||||||:|||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db 101693 CAGTTATATAAAGGGAGGCAGAGCATAAACGTTTGGAAAANTTTCCAGCCTGACAANTGAT 101752

Qy    510 haaraaraaraaycccnttywsngargngnaarttyaarytngcngaratthgyathtg 569
          ||:||:|||||||:||||||||||||||||:|||||||||||||||||:||||:||||:|
Db 101753 AGAAAAGAAAAAACCATTTTTCTGAGGGGGAATTTGAAGCTGCTCAGAAATTTCCATAAG 101812

Qy    570 yaaygargaryytnaaygccncargayaayggngaraayathwsntggacntgyca 629
          ||:|||||||||:||||::|:||||||:|:||||||||||||||||:||||:||||::||
Db 101813 TAACAAGGAGCCTAATGTTAATCCCAAGACAATGGGGAAAANTCTCTCCAAGGCATATCA 101872

Qy    630 rngwsnwsncarcawsnathaarwsnytngcntggmgnaccnnmgnaartggttytg 689
          :|||||||:|||||||||||:||||||||||||||||||||||||:|||||||||||||
Db 101873 GAGGTCTTCATGCAGCCCCATCCACACAGGCCCGGAGCAAAAGGTTTTG 101932

Qy    690 yggnacnggnccnggnwsnytntgtygytgyntncarccnmngtgygt--- 746
          :|||:||||||:|:|:||||||||:|||||||||||:|:|||::|||||
Db 101933 TAGGCCAGGCAGGCCCCATGCTGTGTCAGCCTACGGTGTCCCTGTGTATA 101992

Qy    747 ------------nccngtnaaywsngcngtngcnwsngar 774
                       |||||||||||:|:||||||:|||||:
Db 101993 CCACCTGCTGCAGTGGTGCCTGAAAGGGGCCAGCATAGAGCTTGGGCCCTGGCTTCGGAG 102052

Qy    775 ggngcnwsnccnaarccnctggcarytnccnwsnggngtngarccngtnggngcnaaraar 834
          |||||:||||:::|||||||||:||::::|||:|:||||||||:|||||||||||:|||
Db 102053 GGTGCAACCCCCAAGGCCTTGGCAANTTCCATGTGTTGTTGAGCCTGCACAGAAG 102112

Qy    835 wsnmgnnathaargtntgggarccnccnathmgnttycaaraararattaygynaayccntgg 894
          ||||||:|||||||:||||||||:||||:||||:||||||:|||||||||:||:|||||||:|
Db 102113 TCAAGAATTGAGGCTTGGGAGGCCTCCACCTAGATTTCAGAGGATGTATGGAAAATGCCTTG 102172

Qy    895 atgccnmgncaraaryttggcngtnggngtnsggmgnacnwsngcnmgngtn 954
          ||||||:||||||:|||||||||:|||||||||||||||||||||||||||:
Db 102173 ATGTCCAGGCAGCAGAAGTTTGCTTCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGCA 102232

Qy    955 gtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsn 1014
          ||||||||||:||||||||||||||||||:|||||:|||||||:|:|||||||||||||
Db 102233 GTGCAAAGGGAAAATGTGGGTTGGAGCCCATACACAGAGTCCCTACTGGGGCACCATCT 102292

Qy   1015 wsnmgnmgncgtnmgnmgnwsnccnccnwsnwsnmgnytncaraargngmgnwsnacngay 1074
          |||||||:|:||::||:|||::||||:|||::::||:||||:||||:|||:||||:|||:
Db 102293 AGGGGAGCTGTGAGAGAGGGGCCACCAATCCTTCAGACCCCCAGAATGGTAGAATCCACCAAC 102352

Qy   1075 wsnytncaracaygtnccngaraarwsnacngayacncaargtgycarccngtnaaargcngcn 1134
          ||||:||:|||::|||||:||||:|||||:|||||||||||||:|::||:|||||||:|||::|
Db 102353 AGCTTGCACCATGTACCTGGAAAAGCCACACAGACTCAACACCAGCCCCATAGAAGCAGCC 102412

Qy   1135 ggnatggarwsngtnccntayaaracngtngtncngaryntnacnacacngtnggnath 1194
          |||||||:|:|||||:||||:|||:||||||||:||||:|::||::||||:||||:|:
Db 102413 AAGAGGGAGCCGTGACCCTGCAAAGCCACAGAAGCAGAGCTTCCCAAGGACTTGGGAAC 102472

Qy   1195 tayytnytncaytgyaygayytngaygtnmgncayggngtnaarrmgngayccaytyggn 1254
          ||:||:||:||:||:|:|||:|||||:||:||:||||||||||||:|:||::|||||:|:|
Db 102473 CACCCTTGCATCAGTTGACCTAGATGTGAAACATGCAAAGGAGATCACATATTAGA 102532

Qy   1255 gcnytnmgnttygaytgycnacnggnttymgnacntayatgggngccngt-nccnytntg 1313
          |||||:|::|||||:||:::|:||||||||:|||:|||||||||:|::|::|::||:|||
Db 102533 ACTTTGAGATTTGACTACCCTTCTGGGTTTCAGACTTGCATGGGGCATARGCCCCTTTG 102592

Qy   1314 yttyggncarttyttyccnttyggnaacngcngtntttyacncaartgyytntaayytncaytg 1373
          |||:||||||:|:||||||:||||||||||||||:|:::|:||::||:|:|||::|:|||
Db 102593 TTTTGGCCAAATTTCTCCCATTTAGAACAGCGTGTATTTACTCAATGCCTGTATCGCCATTG 102652

Qy   1374 ya 1375
          :|
Db 102653 CA 102654
```

```
RESULT 3
ACO26107
LOCUS      ACO26107      186660 bp  DNA  linear  HTG 06-FEB-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT
           SEQUENCE, 1 unordered pieces.
ACCESSION  ACO26107
VERSION    ACO26107.22  GI:13899178
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 186660)
AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
           Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
           Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
           Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
           Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
           Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
           Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
           Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
           Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
           Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
           Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
           Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
           Falls,T., Ferraguto,D., Flagg,N., Ford,J.V., Foster,P., Frantz,P.,
           Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
           Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
           Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
           Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
           Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
```

```
                Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
                Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
                Kratovic,J., Kureshi,J., Landry,N., Leal,B., Lewis,L.,
                Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
                Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
                Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
                Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
                Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
                Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
                Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
                Oraguny,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
                Peters,L., Pickens,E., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
                Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
                Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
                Soderberg,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
                Sutton,A., Svatek,T., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
                Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
                Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
                Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
                Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
                Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 186660)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:13877175.
            --------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
            Center project name: HAMK
            Center clone name: RP11-307L1
            --------------- Summary Statistics
            Sequencing vector: Plasmid;
            Sequencing vector: M13;
            Chemistry: Dye-primer BodIpy: 18% of reads
            Chemistry: Dye-terminator Big Dye: 82% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 186630 bases at least Q40
            Consensus quality: 186659 bases at least Q30
            Consensus quality: 186660 bases at least Q20
            Estimated insert size: 187779; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 10.7x in Q20 bases; sum-of-contigs estimation
            ---------------
            *
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 1 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            *   This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            *
FEATURES             Location/Qualifiers
     source          1..186660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-307L1"
BASE COUNT    57089 a  37346 c  35962 g  56263 t
ORIGIN
```

```
Query Match                    Score 497.8; DB 2;  Length 186660;
Best Local Similarity  47.7%;  Pred. No. 1e-108;
Matches  488;  Conservative  216;  Mismatches  282;  Indels  36;  Gaps  2;

Qy    390 rgargaraathwsnaarcaarcarwsnaathcargargtnacntgggtnytnytnaargcntt 449
          :||||:|:|||:|:.:|||::|:.:||.||:| ||::::|.|:|:||:||::|.||::|||:||
Db  71882 GGAAGAAANTTTNCTAAGTNAGTGACAAAGCANTTAAAAGANGGAGGTNGGACTTNAGATGGTNAAAGGTATT 71941

Qy    450 ywsnttyathmgngargtcngarcayaaarwsnwsnqaraaayttncaycccnqayaaygtnat 509
          :|||:|:|.|:||:|||:|:||:|.|:|||:|:||.|||||:|:||:|.:|:.|:|:|||:||:|
Db  71942 CAGTTATATAAGGAAGGCAGAGCANTAAAACGTTTGGGAAAATTTGCAGCCTGACAAATGTGAT 72001

Qy    510 haararaaraaayccnttywsngargngnaarttyaarytngcnqngararathtgyathtg 569
          |:|||||||||:|||:||:|||||:|:|:||:|||||:|:|||.|:||||||||:|:|||:|
Db  72002 AGAAAAAGAAAAAACCATTTTCTGAGGGGAANTTGAAGCTGCTGCAGAAAATTTGCANTAAG 72061

Qy    570 yaaygargaryttnaayccnncargayaayggngararaayathwsntggacntgyca 629
          :|||||:|||:||:|||||:|||:||:|||:||||||||:|||:|||||:|||:||:|
Db  72062 TAACAAGGAGGCCTAANGTTAATCCCAAGACAANTGGGGAAAAATGTCTCCAAGGCANTATCA 72121

Qy    630 rmgnawsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnnaartggttytg 689
          :|||.:|||:||:||:|:|||:||||:|||:||:||.||.|||.|||.|||||:|||||:|:|
Db  72122 GAGGTCTTCANTGGCAGCCCANTCCAATCACAGGGCCCGAGGCCCTAGGAGAGAAATTGGTTTTG 72181

Qy    690 yggnacnggngnccngnmwsnytntngtygtgtncaarccnmngayyttngtncccntgygt--- 746
          :||||:|||:|:|||:|:.|||:|:|:|:|:|:||||||:|:|||:|:||:||:|||:||:||
Db  72182 TAGGCCAGGCCCCAGGCACCCCAGGACCCCATGCTGTCTCCAGCCTAGGGGACTTGGTGCCCTGTGTATA 72241

Qy    747 -----------nccngtnaaywsngcngtngcnwsngar 774
                      :|||:||:|||:|||:|:|||||:||:||:|||
Db  72242 CCAGCTGCTGCAGTGGTGGCTGAAAAGGGCCAGCATNAGAGCGTTGGGCCGTGGCTTCGGAG 72301

Qy    775 ggngncnwsnccnaaarccnctggcaarytgnccnwsnggngtngarccngtngggngcnaaraar 834
          |||:|:|:|||||||||||:|||||||:|||:|||.|||||:|||||||||:|||:|:|||||||:|
Db  72302 GGTGCANGCCCCAAGCCCTTGGCANTTCCATGTGGTGTTGAGCCTGGAGGGTGCACAGAAG 72361

Qy    835 wsnmgnathgargtntgggarcncnccnaathmgnttycaaraaraathtayggnaayccntgg 894
          .|||:||:|||||:|:||||:|:|:|||:|||:||:||:||||||||:|||:|||:|||:|||||
Db  72362 TCAAGAANTTGAGGCTTGGGAGCCTCCACCTAGATTTCAGAGGATNGTATGGAAAATGCCTGG 72421

Qy    895 atgcccnmgncaraaarttygncngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtn 954
          |||||:|.|||:|||||:||:|:|:|:||:|:||:|.|||.|||||.||:.|||:|:|:||:|:
Db  72422 ATGTCCAGGCACAGAAGTTTNGCTTCAGGGGTGCCCCATGGAGNAGGAACCCTCTGCTAGGGCA 72481

Qy    955 gtncaraaargngnaaaygtnaggnttgggggarcncnccncaymgngtncccnwsnggngcnccnwsn 1014
          |:|||||||||:|:|||||:|||.|:||:|:||:|:|:|||:|||:|||:||||.|||||||:|:|||.|||
Db  72482 GTGCAGAAGGGAAAATGTGGGGTTGGAGCCCATAACACAGAGTCCCTACTGGGGCACCATCT 72541

Qy   1015 wsnmgngncngtnmgmgngnwsnccncccnwswsnmgnytncaaraargngnmgnwsnacngay 1074
          .|||:|||:|:||:|.|:|:|.||||||||.|.|||:|:||:||||||||:|:|||.|||||:|||
Db  72542 AGGGGACCTGTGAAGAGAGGGCCCACCAGAAATGGTAGATCCACCAAC 72601

Qy   1075 wsnytncarcaygtnccngaraarwsnacngayacncaartgycarccngtnaaargcngcn 1134
          .|||:||||:||:||:|||||||:.|||||:||||:|||:|||:|||||||:||||||:|||:|
Db  72602 AGCTTGCACCATGTACCTGGAAAAGCCACACAGACTCAAACACCACCATGAAAAGCAGCC 72661

Qy   1135 ggnatggarwsngtnccntayaaaracngtngtncngaarytnacnaaracngtngggnath 1194
          ||:|||||:|||:||:|||:||:|||||:||:|:|:|||||:||||:||||||:||||||:||:
Db  72662 AAGAGGGAGCCGCTTACCCTGCAAAGCCACAGAAGCAGAGCTTCCCAAGACACCATGGGGAACC 72721

Qy   1195 tayytnytncaytgycaygayytngaygtnmgncaygggngtnaaarmgngaycaytyggn 1254
          |||:||:|:|||:||:|||||:||:|||||:|:|||||||:|:|||||:|:|||:|:|:|||
Db  72722 CACCTCTTGCACCATGTGACCTAGATGTGAAACATGCCAGTCAAAAGGAGATCATATTAGA 72781

Qy   1255 gcnytnmgntygaytgyccnacngngnttymgncacntayatgggnccngt-nccnytntg 1313
          |||:||:|||:||:||:|||||:|:|:||:|||:|||:||:|||||:|||:| :|||:|:||
Db  72782 ACTTTGAGATTTGACTNACCCTNCTCTGGGTTTCAGACTTGCATGGGCCATNAGCCCCTTTG 72841

Qy   1314 yttyggncaarttyttyccnttyggnacncgngtnttyacnacaartgyyttntayytncaytg 1373
          :||:||:|||:||:||:|||:||||:|||:|:|||:||:||:||||||:|:||:||:|:|||:||
Db  72842 TTTTTGGCCAATTTNCTCCCATTTAGAACAAGCTGTTATTTACTCAATGGCCTGTATCGGCCCATTG 72901

Qy   1374 ya 1375
```

```
Db  72902 CA 72903
            ::|

RESULT   4

LOCUS       AL138962/c

DEFINITION  AL138962   12418 bp    DNA    linear    PRI 28-JAN-2001
            Human DNA sequence from clone RP11-279N8 on chromosome 13, complete
            sequence.
ACCESSION   AL138962
VERSION     AL138962.16  GI:12597036
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124518)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 29, 2001 this sequence version replaced gi:12584354.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-279N8 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-279N8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-279N8 is at 124518 in this
            sequence. The true left end of clone RP11-528024 is at 114969 in
            this sequence. The true right end of clone RP11-214011 is at 100 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1..124518
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-279N8"
                     /clone_lib="RPCI-11.1"
     misc_feature    112388..112496
                     /note="Big dye primer and dGTP big dye chemistries only."
BASE COUNT    39295 a   26385 c   24794 g   34044 t
ORIGIN
```

```
Qy    390 rgargarathwsnaarcarcaawsnathcaargargtnacttgggtnytnaaragcntt 449
          ::||:||  |||:||  |   ||:: ||||:|| || |||||:|| ::| ||:|| |
Query Match    35.8%;  Score 493.6;  DB 9;  Length 124518;
Best Local Similarity  47.5%;  Pred. No. 9.2e-108;
Matches 486;  Conservative 216;  Mismatches 285;  Indels 36;  Gaps 2;
```

```
Db  66716 GGAAGACATTTCTTAGCCAGAGAAAAAGCATTCAAGAGAGGTGACTTGGGTACTGTTPAGAGGCATC 66657
          :||  | |   | |    |   |||    |:| ::| ||:|| || | |  ::|  ||| |
Qy    450 ywsnttyathmngargcngarcayaarwsnwsngararaayytncaycccngayaaygtnat 509
          ::| |:||  | :|| |:|| ||| || || || | ::|| | ||:| |:||||| :|||
Db  66656 CAGTTTATAAGGGAAACAGAGTATAAAAGTTGAAAAATTTGCAGCCTGACAAATGTGAT 66597
          |:|| |||||:||  | ||| | ||||:|| |||||| ||:| || ||  ||| |||
Qy    510 haaraaaraaraaayccnttywsngargggnaarttyaarytngcngcngarathtgyathtg 569
          |||||||||||||||  |:|| | | | |||||| ||:|| | | | | ||| | |||
Db  66596 AGAAAAGAAAAATCCCANTTTCTGAGGAGAAAGTCAAGCTGCTGCAGAAATTTGCATAAG 66537
          | | |:|| ||||| | || || |||| ||||||||| ||||||| |||||| ||| ||
Qy    570 yaaygargaryytnaaydytnaaycacargayaaydgggagaraayathwsntggacntgyca 629
          :|| ||| || :|||:|  | ||||||||   ||:|| || || |:||| | || || ||
Db  66536 TAACGAGGACGCCAAAATGTTAATCACCAAGACAATGGAGGAAATGTCTCCAGGGCATGTCA 66477
          ||||||| || |||||||||| |||| || ||||| || || ||||  || |||||| |
Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg----naatgtgtt 686
          |:|| || || || || ||| | ||||||||:|| | ::||  |:|:      || ||||
Db  66476 GAGGTCATCACAGCATCCCCTCCAAGACACGCCTGGAAGCTTAGGAGGAAAAAATGGTT 66417
          |||||||||| || |||| |: | |||||| |  ||| |  |::||| ||||| ||||
Qy    687 ytgyggnacnggngncconggnwsnaytcntgytgyggtncarccnmgnayytngtnccntgyg 746
          :||:|| || |||  |   |||:|  |:||||:||:||||  |  |:|| ::||:||||| |
Db  66416 TCATGGAAGGGGCACAGGGTCCCTGYTTGYTGTGCAGCCTAGGGGGACTTGGTGTCCTGCAT 66357
          |||||| || || | ||||  || |::|| :|| ||| ||| ||||| ||| |||| ||
Qy    747 nccngtnaayws----------------ngcngtngcnwsnga 773
          |:| || ||:                  ||||||:|| ||
Db  66356 CCCAGCTGCTCCAGCTGTGCCAGCTGTGCCAAATGTGGGAGGTTCAGGCTGTGGCTTCAGA 66297
          ||| || || |||  |||  || ||  || | |||| || ||||||| ||| | |||| |
Qy    774 rggngcnwsnccnaacctnggcaryytnccnwsngggntngarccngtnggngcnaaraa 833
          |:|||  | || || |||||   :|:|| |:|||:||:|| || | ||||| | ||||
Db  66296 GGGTGCAAGCCCTAAGCCTNGGCAGTTCCACATGATGTTAAGGCTGTGAGTGCACAGAA 66237
          ||||||||||||  ||||:|| ||||||||||||||||| || ||||||||| ||||
Qy    834 rwsnsnngnathgargtntggarccncccnathmgnttycararathtaygnaayccntg 893
          |||:|:| | ||| |||  |||||| | | |  |:||||| || |:|| |||| | |:
Db  66236 ATCAAGAACTGGGTGTCAGAGTTCATCAGATAATGTTAATACCTG 66177
          || ||||| ||| |||| ||||||||| || | ||||| || |
Qy    894 gatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngt 953
          | ||||| | | || ||||| |||||:||:|||:|| || ||| || || | |  |:||
Db  66176 GATGCCAGACAGACAGAAGTTTGCTGTAPGGGAGGGGTGTCATGAAAAACCTCTGCTAGGGC 66117
          | ||||| || || ||||||||| ||||  |||||||||||  |||||| || || |||
Qy    954 ngtncaraarggnaaytnggntnggnagarcnccncacaymgngtnccnwsnggngncccnws 1013
          ||:|| || |||||||||:||:|||:|  || | | | |:|||:||| |||:||:|| | |
Db  66116 AGTGTGGGAGGGGAAAATGTGGGCCCATTCCACAGAGTCCCTACTGGGGCACCGCC 66057
          |||||||||||||||||||||||| |  |:||||| || |  ||||| ||| ||
Qy   1014 nwsnsnmgncngtnmgmgnwsncccnccnwsnsnmgnytncararargggnmgnwsnacnga 1073
          ||:|:|:|| |||:|:||:||| || ||  |:|:|| :||| | |||||:|| || || |
Db  66056 TAGTGGAGCTGTGAAAAGTGGGCCACTATCCTCCAGCACCTGAGATCCACTGA 65997
          ||||| ||||||||||||||||| || | |||| || | ||| | ||| ||
Qy   1074 ywsnnytncarcaygtnccngararawsnacngayacncartgycarcccngtnaargcngc 1133
          :|||:|||||| || || |||| | || || || |:|| ||  | || ||| |||| |||
Db  65996 CAGCTGTGCACCATGCACCTGGAAAAGCCAGAGACACTCAAATGCCAGCCTGTGAAAGCAGC 65937
          ||| ||| ||| ||| || |||||||| || ||| | || || || || || |||| |||
Qy   1134 nggnatggarwsngtnccntayaaracngtngtngngaryytnacnaaracngtnggnat 1193
          |||||||| | ||:|| |||||||| ||:||:||:|| :||| |||||||||:|||| |
Db  65936 TGGGAGGGGAGACTGTACCATGCAAGACCACAGGGGTGCCCAAGACCCAATGGGAAC 65877
          ||||| || || || |||||| || ||||||:||:|| ||||| |||||||||||
Qy   1194 htayytnytncaytgycaygaaytngaygtnmgncaygggntnaarmgngaaycaytygg 1253
          ||| :||:|||| || |||| ||:| ||:|:|| || ||:|||| |:|||| |||:||
Db  65876 CCACCTGTTGCATCAGCATGGATGTGACANTGGAGGAGATCAGTTTTGG 65817
          ||| ||:||||| | |||| || |||| ||||||||| | |||||||
Qy   1254 ngcnytnmgnttygaytgyccnacnggnqntymgnacntayatgggnaccngtnccnytntg 1313
          ||| :||:|| :|| |||   || || ::|||:|| || || ||| |||:||||| :|:||
Db  65816 AGCTTTAAGATTTGACTTCCCTGCTGGATTTCAGACTTGCAATGGGCCCTTG 65757
          |||| || | ||||||| ||| || ||| |||||| || |||||| |||||
Qy   1314 yttyggncaartyttyccnttyggnacngcngtntyacncaartgyytnayytncaytg 1373
          :||:|| || || |||||:||:||||| || ||:|| || || || :||||:|||||:
Db  65756 TTTGGCCAATTCCCCCATTGGAATGGGGTATAPTACCCAATGCCTGTACCCCATTG 65697
          |||| ||||| ||||||||||||| || ||| |:||||||||| || ||||||||
Qy   1374 yat 1376
          :||
Db  65696 TAT 65694
```

```
LOCUS       AC008799              123070 bp    DNA     linear   HTG 18-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2061E19, WORKING DRAFT
            SEQUENCE, 8 ordered pieces.
ACCESSION   AC008799
VERSION     AC008799.4  GI:9256046
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7709316.
            -----------Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
            ----------
            Project Information
              Center Project Name: 651088
              Center clone name: CITB-H1_2061E19
            ----------
            Summary Statistics
              Consensus quality: 116295 bases at least Q40
              Consensus quality: 121288 bases at least Q30
              Consensus quality: 122086 bases at least Q20
              Estimated insert size: 123000; pulse-field gel estimation
              Estimated insert size: 122770; sum-of-contigs estimation
              Quality coverage: 6.04 in Q20 bases; pulse field gel estimation
              Quality coverage: 6.05 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     9436: contig of 9436 bp in length
            *     9437     9536: gap of unknown length
            *     9537    26537: contig of 17001 bp in length
            *    26538    26637: gap of unknown length
            *    26638    45958: contig of 19321 bp in length
            *    45959    46058: gap of unknown length
            *    46059    68657: contig of 22599 bp in length
            *    68658    68757: gap of unknown length
            *    68758    72152: contig of 3395 bp in length
            *    72153    72252: gap of unknown length
            *    72253    81069: contig of 8817 bp in length
            *    81070    81169: gap of unknown length
            *    81170   121547: contig of 40378 bp in length
            *   121548   121647: gap of unknown length
            *   121648   123070: contig of 1423 bp in length
FEATURES             Location/Qualifiers
     source          1..123070
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone_lib="CalTech human BAC library D"
                     /clone="CTD-2061E19"
BASE COUNT    37763 a  24223 c  23855 g  36528 t    701 others
ORIGIN
```

Query Match   35.8%;   Score 493;   DB 2;   Length 123070;

Best Local Similarity 46.7%;   Pred. No. 1.3e-107;
Matches 495; Conservative 220; Mismatches 309; Indels 36; Gaps 2;

```
Qy    352 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 411
          ||| |  |||  |||  : | || | |:  ||||   |||| : |:
Db 107907 AACTTTTAACTTGAAAGAAATGATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 107966

Qy    412 wsnathcaargtnacntcgggtnytnytnaargcnttywsnttyathmgnaargcngar 471
          :| :| || ||  |  ||||| |||| |  |||||  |: | | | ||   ||| | ||
Db 107967 AGCATTCCAGAGGTGACTTGGGTCGTGTTGTAAAGGTATTCAGTTTTATAAGGAAGCAGAG 108026

Qy    472 cayaarwsnwsngaraayytncaycongayaaygtnathaaraaraaaaaycnttywsn 531
          || ||| :| :|| ||| |||:  |:||  || || |  ||||| || ||| |: | :|
Db 108027 CATAAAAGTCTTGGAAAATTTGCAGTCCGACAAATGTGATAGAGAAAAGAAAATTCCATTTTCT 108086

Qy    532 gargnaarttyaarytngcngcngarathtgyathwsngacntgycaryngmgnwsnwsncarcaarwsnath 591
          || | || || || ||| || ||||  || |  :| || | ||  ||  |  || | : | | || |  |
Db 108087 GAGGAGAAATTCAAGCTGGCTGGCAGAAATGTGCATAAGTAACAACAGAGCTGAAATATTAAT 108146

Qy    592 ccncargayaaygngaraayathwsntggacntgycaryngnwsnwsncarcaarwsnath 651
          |:| || ||  |    |||| |  :| || |||  ||  |  || |  | | ||  :| |
Db 108147 CCCCAAGACAAATGGGGAAAAACATCTCCAAGGCATGTCAGAGGTCTTCACGGACGCCCTC 108206

Qy    652 aarwsnaytngcntggmgncnmgnmgng- - -naartggtttygygngnacngnccnggnwsn 708
          :|: || |||  || | | | ||| |    ||| || |||| | | |  |||| |||| :|
Db 108207 CCATCATAAACCCAAAGCCCTAGGAGGAGAAAAAAGTGGTTTTGTGGGGCCAGGCCCAGGGTCC 108266

Qy    709 ytntgytgygtncarccnmgnagytngtnccntgyygttnccngtnaa- - - - - - - - - 755
          :|:  | |||| | || || |  | |||| |||| | | || || ||
Db 108267 TTGTGCTGTGTGTCGCAGCTTAGGGACTTGGTGCTGGCGCTGCTTCAGCTGCTC 108326

Qy    756 - - - - - - - - - ywsngcngtngcnwsnaargngngcnwsnccnaarccntgg 795
                     |  |: | |||| | | | | || ||  ||| |||:|
Db 108327 GAAAGGGGCCAACATAGAGCTCGGGCTCGTGGCCTTCAGAAGGTGCAAGCCCCAAGCCTTGG 108386

Qy    796 carytnccnwsnggnggtngarcccngtnggngcnaaraarwsnmgnahgargtnggngar 855
          || | |:| :| || || || || |||| || |||| || | || ||  |  | | | ||
Db 108387 CAGCTTCCATGTGGTCGTTGTGAGCGTGCAAGTGCACAGAGTCACAAAATTGGGGTTTGGGAA 108446

Qy    856 ccncncnaathmgnttycaararathtaygngnaayccntggatgccnmgncaraarttygcn 915
          |:| |||||  |||| ||| | |  |  || | |||||  |||  ||  |||| | |||| |
Db 108447 CCTCCACCTAGATTTGCAGATGATGTTATGGAAAATGCCCGGATGACCCAGGACAGAAGTTTGCT 108506

Qy    916 gtnggngtnggnwsnsntggmgncacnwsngcngtngtncaraargngaaygtnggn 975
          || || || || :| :| || | | | :| || || || | ||| | | | || ||
Db 108507 GCAGGGGTGAGGGCCGTTCATGGAGGAACCTCCTAGGGCAGTGCGAAGGAGAAAAATGTAGGG 108566

Qy    976 tgggarccncnccnacaymgnggtnccnwsnsgncgncnwsnsnmgncngngtnmgngnwsn 1035
          || || |:| |  || | || || ||| :| :| | | | |:| || | || |:| | :|
Db 108567 TCAGAGCCCACACAGAGCTCGGGCTCGTGGCCTTCAGAAGGTGCAAGCCCCAAGCCTTGG 108626

Qy   1036 ccnccnwsnswsnmgnytncaraarggnmgnwsnacngaywsnytncaarcaygtnccngar 1095
          |:| ::| :| :|  || | | || ||   | :| || | :| | || || |||| ||| |
Db 108627 CCACCATCCTCCAGACTCCAGAAATGTTAGGGTCCAAACAGCTTCGACTGTGTACCTGGA 108686

Qy   1096 aarwsnacngaryacncaartgycarccngtnaargcngcngnatggarwsngtnccntay 1155
          || :| || ||  | | || |||| || || || || || | || || :| || |||| |
Db 108687 AAAGCCACAGGCCACTCATTGCCAGCCTGTGAAAGCAGCCAGGAGAGGGCTATCCCCTGC 108746

Qy   1156 aaracngtngcngaryacnaaracngtngnathtayytnytncaytgtgycaygay 1215
          || || || || ||  | ||  | || |  | ||| | | || ||| | | || |
Db 108747 AAAGTCACAGGCGGCGAAGCCTGTCCAAGAACCTCTTGCCCTCAGCCTGAC 108806

Qy   1216 ytngaygtnmgncayggngtnaaramgnaycaytyggngcnytnmgnnttygaytgycn 1275
          | |  | || || ||  | |||  |   | |: |:| |  || || |  |  |  | :|
Db 108807 CCGGATGTGACACATGGAGTCAAAGAGCATCATTTGGAGCTTTAAGATTTGACTGCCT 108866

Qy   1276 acnggnttymgnacntayatggggnccngtnccnytngtttyggncaarttyytccnty 1335
          || || ||: || || |  || ||:| || |||: |  ||| || | || | | || :|
Db 108867 GCTGGATTTCAGACTGCATGGGGGCTGTGTAGCCCCTGCTTTGGCCAAATTTCTCCCATTT 108926

Qy   1336 ggnacngcngtnttyacncaartgyytntayytncaytgya 1375
          || || | || ||  || || |  || | | ||  ||: |
Db         (continues)
```

Db 108927 GGAACAGCTGTATTTACCCATGCCTGTCTGTACCCGATTGTA 108966

```
RESULT    6
ACC044889/c
LOCUS          AC044889      189768 bp    DNA      linear    HTG 22-MAY-2000
DEFINITION     Homo sapiens chromosome 2 clone RP11-792C1 map 2, WORKING DRAFT
               SEQUENCE, 36 unordered pieces.
ACCESSION      AC044889
VERSION        AC044889.2  GI:8016676
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 189768)
AUTHORS        Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE          Homo sapiens chromosome 2, clone RP11-792C1
JOURNAL        Unpublished
REFERENCE      2  (bases 1 to 189768)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
               Boguslavkiy,L., Bukhgalter,B., Brown,A., Burkett,G.,
               Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
               Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
               Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
               Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
               Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
               Klein,J., Lakocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
               Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
               McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
               Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
               Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
               Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
               Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Theodore,J., Tirrelli,A., Travers,M., Triglio,J.,
               Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
               Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        On May 22, 2000 this sequence version replaced gi:7543856.
               All repeats were identified using RepeatMasker:
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ---------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ---------------- Project Information
               Center project name: L5594
               Center clone name: 792_C_1
               ---------------- Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 168749 bases at least Q40
               Consensus quality: 180047 bases at least Q30
               Consensus quality: 184014 bases at least Q20
               Insert size: 190000; agarose-fp
               Insert size: 186268; sum-of-contigs
               Quality coverage: 3.8 in Q20 bases; agarose-fp
               Quality coverage: 3.9 in Q20 bases; sum-of-contigs
               ----------------
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 36 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*           1  1015: contig of 1015 bp in length
*        1016  1115: gap of  100 bp
*        1116  2284: contig of 1169 bp in length
*        2285  2384: gap of  100 bp
*        2385  3898: contig of 1514 bp in length
*        3899  3998: gap of  100 bp
*        3999  5259: contig of 1261 bp in length
*        5260  5359: gap of  100 bp
*        5360  6912: contig of 1553 bp in length
*        6913  7012: gap of  100 bp
*        7013  8145: contig of 1133 bp in length
*        8146  8245: gap of  100 bp
*        8246  9694: contig of 1449 bp in length
*        9695  9794: gap of  100 bp
*        9795 11004: contig of 1210 bp in length
*       11005 11104: gap of  100 bp
*       11105 12907: contig of 1803 bp in length
*       12908 13007: gap of  100 bp
*       13008 14916: contig of 1909 bp in length
*       14917 15016: gap of  100 bp
*       15017 17114: contig of 2098 bp in length
*       17115 17214: gap of  100 bp
*       17215 18527: contig of 1313 bp in length
*       18528 18627: gap of  100 bp
*       18628 20561: contig of 1934 bp in length
*       20562 20661: gap of  100 bp
*       20662 22484: contig of 1823 bp in length
*       22485 22584: gap of  100 bp
*       22585 25287: contig of 2703 bp in length
*       25288 25387: gap of  100 bp
*       25388 27603: contig of 2216 bp in length
*       27604 27703: gap of  100 bp
*       27704 30020: contig of 2317 bp in length
*       30021 30120: gap of  100 bp
*       30121 33472: contig of 3352 bp in length
*       33473 33572: gap of  100 bp
*       33573 38084: contig of 4512 bp in length
*       38085 38184: gap of  100 bp
*       38185 41069: contig of 2885 bp in length
*       41070 41169: gap of  100 bp
*       41170 45266: contig of 4097 bp in length
*       45267 45366: gap of  100 bp
*       45367 51092: contig of 5726 bp in length
*       51093 51192: gap of  100 bp
*       51193 55262: contig of 4070 bp in length
*       55263 55362: gap of  100 bp
*       55363 60982: contig of 5620 bp in length
*       60983 61082: gap of  100 bp
*       61083 67628: contig of 6546 bp in length
*       67629 67728: gap of  100 bp
*       67729 72690: contig of 4962 bp in length
*       72691 72790: gap of  100 bp
*       72791 79286: contig of 6496 bp in length
*       79287 79386: gap of  100 bp
*       79387 83451: contig of 4065 bp in length
*       83452 83551: gap of  100 bp
*       83552 92170: contig of 8619 bp in length
*       92171 92270: gap of  100 bp
*       92271 96333: contig of 4063 bp in length
*       96334 96433: gap of  100 bp
*       96434 96433: gap of  100 bp
*       96434  9643: gap of  100 bp
*       96434 103218: contig of 6785 bp in length
*      103219 103318: gap of  100 bp
*      103319 112553: contig of 9235 bp in length
*      112554 112653: gap of  100 bp
*      112654 123239: contig of 10586 bp in length
*      123240 123339: gap of  100 bp
*      123340 137921: contig of 14582 bp in length
*      137922 138021: gap of  100 bp
*      138022 157913: contig of 19892 bp in length
*      157914 158013: gap of  100 bp
```

```
FEATURES    Location/Qualifiers
  *         158014   189768: contig of 31755 bp in length.
  source    1..189768
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2"
            /clone="RP11-792C1"
            /clone_lib="RPCI-11 Human Male BAC"
  misc_feature  1..1015
            /note="assembly_fragment"
  misc_feature  1116..2284
            /note="assembly_fragment"
  misc_feature  2385..3698
            /note="assembly_fragment"
  misc_feature  3999..5259
            /note="assembly_fragment"
  misc_feature  5360..6912
            /note="assembly_fragment"
  misc_feature  7013..8145
            /note="assembly_fragment"
  misc_feature  8246..9694
            /note="assembly_fragment"
  misc_feature  9795..11004
            /note="assembly_fragment"
  misc_feature  11105..12907
            /note="assembly_fragment"
  misc_feature  13008..14916
            /note="assembly_fragment"
  misc_feature  15017..17114
            /note="assembly_fragment"
  misc_feature  17215..18527
            /note="assembly_fragment"
  misc_feature  18628..20561
            /note="assembly_fragment"
  misc_feature  20662..22484
            /note="assembly_fragment"
  misc_feature  22585..25287
            /note="assembly_fragment"
  misc_feature  25388..27603
            /note="assembly_fragment"
  misc_feature  27704..30020
            /note="assembly_fragment"
  misc_feature  30121..33472
            /note="assembly_fragment"
  misc_feature  33573..38084
            /note="assembly_fragment"
  misc_feature  38185..41069
            /note="assembly_fragment"
  misc_feature  41170..45266
            /note="assembly_fragment"
  misc_feature  45367..51092
            /note="assembly_fragment"
            /note="assembly_fragment"
            /vector_side:right"
            /clone_end:T7
  misc_feature  51193..5262
            /note="assembly_fragment"
  misc_feature  5363..60982
            /note="assembly_fragment"

Query Match      35.8%;    Score 493;  DB 2;  Length 189768;
Best Local Similarity  46.7%;   Pred. No. 1.5e-107;
Matches   495;  Conservative 220;  Mismatches 309;  Indels   36;  Gaps   2;

Qy    352 aayathggytnatgngnaarcarathytngcnaayaargargarathwsnaarcarcar 411
          ||:|.|.||                      :||||.|.||  |||  ||||.||.||||::
Db  86595 AACTTTTAACTTGAAAGAAATGATTTAGGGTATCTGGCAGAGAGAAAATTTCTAAGCACCAA 86536

Qy    412 wsnathcargargtnacntgggtnyttnaargncnttywsnttyathmgngarngcngar 471
          ::|:|||:|||:|:||||:|:|:|||||.||  ||:|||||.|.||  |:||.|:|.||
Db  86535 AGCATTCCAAGAGTGACTTGGGTGCTGTTTAAAGGTATTCAGTTTTTATTTAPAAGGAAGCAGAG 86476

Qy    472 cayaarwswsngaraayytncayccngayaaygtnathaaraaaraayccnttywsn 531
```

```
Db  86475 CATAAAAGTCTGAAAAAATTTGCAGTCCGACAAGTGAYGAGAAAAGAAAAATTCATTTTCT 86416
          ||:||.|::

Qy    532 gargggnaartyaaarytngcngngaarathtgyathgyaaysargargrytnaaygtnaay 591
          |:|:.|:||  |||| || ||  ||||.:||:||||:|||.::.||:||.||:||||.||
Db  86415 GAGGAGAGAAATTCAAGCYGCTGGCTGCGAGAAAATGTGCATAAGTAACAAAGGAGCTGAATATTAAT 86356

Qy    592 cncacngayaaygngaaraayaaayatathwsntcggacntgycarmgnwsnwsncarcarwsnath 651
          ||.:|.|||:||||:|||||:|::|:.||:|||.:||  |||||||.||.:||||  |||.|:|
Db  86355 CCCAAGACAAATGGGAAAAACATCTCCAAGGCAATGTCAGAGGGTCTTCACGGAGCCCCTC 86296

Qy    709 ytntgytgygtncarccnmgnaayytgtnccntgyytgyggtnccngtnaa--------- 755
          :||||||||:||||  ||.:||||.|:||:.||.::|.||||||:.||||
Db  86295 TTGTGCTGTGTGCAGACTTAGGGACTTGGTGCCCTGCCTTCAGAGGGTCC 86236

Qy    756 ----------------ywsngcngtngcnwsngargnggngcnwsnccnaarccntgg 795
                          :||.||.||||||||.||||.|.||.||||:|||.|:.:||
Db  86235 GAAAGGGGCCAACATAGCAGCTCGGCGCTGGCTTCAGAAGGGTGCAAGCCCCAAGCCTTGG 86176

Qy    796 carytnccnwsngggngtngarccngtnggngcnaaraaarwsnmgnathgargtntgggar 855
          :||: ||||.||||||||||||.|||||.||:.||||:||||.|.|||||.||||||||
Db  86175 CAGCTTCCATGTGTGGTGTTGAGCCTGCAAGTGCACAAGAAGTCACAAAATTGGGGTTTGGGAA 86116

Qy    856 ccnccnathmgnttycaraaarathtaygngnaayccntggatgccnmgncaraarttygcn 915
          |:|.:|.||:||||  |||.|||.:||.|||.||||.|:.|||||:||||.|||||||.|:|
Db  86115 CAGCTTCCATGTGTGGTGTTGAGCCTGCAAGTGCACAAGAAGTCACAAAATTGGGAA 86056

Qy    916 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncararggnaaygtnggn 975
          |||||||.||||||  ||.||.|||||:|.||.:|.||:.||||  |||.::||||||
Db  86055 GCAGGGGTGCAGGGCCTTCATGGAGAACCCTCTGAGGCAGTGCAGAAGAAGGTTCGCT 85996

Qy    976 tgggaarccncncncaymgnmgytncnncaymgncaccnwsnwsnmgncngtnmgnmgnwsn 1035
          |||.:|||.|||.|.:|.|||||  ||.::||.|.:||.:|||.:|.||.:.|:.|.||
Db  85995 TCAGAGCGCCACACACAGACTACCCACTAGAGCACTGCCTAGTGAGCCTGTGAGAAGAAGG 85936

Qy   1036 ccncnwsnwsnmgnytncaraaraarggnmgnwsnacngayawsnnytncarcaygtnccngar 1095
          |:||.::|||  ||.||.||.|||  |||:||||  ||.||.:|||||.||||  ||.|||.||
Db  85935 CCACCATCCTCCAGACTCCAGAATGGTAGGTCCACCAACAGCTTGCACTGTGTACCTGGA 85876

Qy   1096 aarwsnacngayacnaartgycarccngtnaargcngngnatggarwsngtnccntay 1155
          |||.::|||||.||||||  |.||.:||||||.||.|.||||  :||.||||.||||
Db  85875 AAAGCCACAGGCACACTCATTGCCACCTGTGAAACCAGCCGAGGAGGGCTATCCCCTGC 85816

Qy   1156 aaracngtngtncngaryytnacnaaracngtnggnathtaytytnytncaytgycaygay 1215
          ||:|  |||.||.||||.:||.:|.|||.|||||.||.:|.||||.:|.||.|.||||||||
Db  85815 AAAGCCACAGGCACACTCATTGCCACCTGTGAAACCAGCCGAGGAGGGCTATCCCCTGC 85756

Qy   1216 ytngaygtnmgncayggngtnaarmgnsgaycayttyggngcnytnmgnttyggaytgycn 1275
          :|:||:|||||:||||.|||.|||||.|.|.|||||:|||||||.|||.||||.||||||
Db  85755 AAACTCACAGGGGCGGAAGCTGTCCAAGACCTTGGAACCTACCGTGAC 85696

Qy   1276 acnggnttymgnaacntayatgggngcnccngtncnytntgtyttyggncaartytyccntty 1335
          |.|| |||.|:  ||||||.||||||:|||.|:||||.:||:||.|||||.||||||.||||
Db  85695 CCGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCT 85636

Qy   1336 ggnacngcngtnttyacncaartgyytntcaytayytgya 1375
          :|||.||:|||||||||||||||.:|:||||:|||| :|||:|
Db  85635 GCTGGATTTCAGACGTGCATAGGGGGCCCCTTTGCTTTGGCCCAATTTCTCCCATTT 85576

Db  85575 GGAACAGCTGTATTTTACCCCATGCCTGTACCCGAATTGTA 85536
```

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 198545)
  AUTHORS  Waterston,R.H.
  TITLE    The sequence of Homo sapiens clone
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 198545)
  AUTHORS  Waterston,R.H.
  TITLE    Direct Submission
  JOURNAL  Submitted (09-DEC-1999) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
COMMENT    On Nov 2, 2001 this sequence version replaced gi:15741601.

           --------------- Genome Center ---------------
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu/gsc/index.shtml
           Contact: submissions@watson.wustl.edu
           ------------ Project Information ------------
           Center project name: H_NH0354H17
           ---------------- Summary Statistics ----------------
           Sequencing vector: M13; 35%
           Sequencing vector: plasmid; 65%
           Chemistry: Dye-primer ET; 34% of reads
           Chemistry: Dye-terminator Big Dye; 66% of reads
           Assembly program: Phrap; version 0.990319
           Consensus quality: 197437 bases at least Q40
           Consensus quality: 198145 bases at least Q30
           Consensus quality: 198648 bases at least Q20
           Insert size: 200000; agarose-fp
           Insert size: 200801; sum-of-contigs
           Quality coverage: 10.58 in Q20 bases; agarose-fp
           Quality coverage: 10.55 in Q20 bases; sum-of-contigs

         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 4 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *
         *       1            1174: contig of 1174 bp in length
         *       1175         1274: gap of unknown length
         *       1275         2716: contig of 1442 bp in length
         *       2717         2816: gap of unknown length
         *       2817         4952: contig of 2136 bp in length
         *       4953         5052: gap of unknown length
         *       5053       198545: contig of 193493 bp in length.
FEATURES             Location/Qualifiers
     source          1..198545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-354H17"
     misc_feature    1..1174
                     /note="assembly_name:Contig24"
     misc_feature    1275..2716
                     /note="assembly_name:Contig38"
     misc_feature    2817..4952
                     /note="assembly_name:Contig42"
     misc_feature    5053..198545
                     /note="assembly_name:Contig44
                     vector_side:right"
                     clone_end:SP6
BASE COUNT     62963 a  36710 c  37248 g  61324 t    300 others
ORIGIN
```

Query Match   35.6%;   Score 490.4;   DB 2;   Length 198545;

Best Local Similarity   47.0%;   Pred. No. 6.5e-107;
Matches 498;   Conservative 213;   Mismatches 314;   Indels 34;   Gaps 2;

```
Qy    352 aayathtggytnaatgmgnaaarcaarathytngcnaayaaargargaarthwsnaarcarcar 411
          ||:|  |  |:|:  ||| | | |  | | :| : | |: |:| |  |:|| :::|:|||
Db  83103 AACTTTTAACTTGGGAGAGAGATCGATTTTAGGGTGATGATCTGGCAGAATAAATTTCTAAGCAGCAA 83162

Qy    412 wsnatheaargargtnacntgggtnytnytnaaargcnttywsnttyathmgnaargcngar 471
          :: ||:|:||:||:|| |||| || | :|| | | || :| || |: ||||| |:|
Db  83163 AGCATTCAAGCACGTGACTTGGGTGCTGTTAAAGGCANTTCAGTTTTATAAGGGAAAGCAGAG 83222

Qy    472 cayaayswsnsngaraaayytncaycayccngayaaaygtnathaaraaaraaaaaycnttywsn 531
          || :||:||  |  | ||:|| || ||| | | |:::||:  ||:|| |:| ||||||  |
Db  83223 CATAAAAAATTTGGAAAAATTTTGCAGCCTGACACAATGTGATAGAAAAAGAAAAAACCCATTTTCT 83282

Qy    532 garggnaaartyaaarytngcngncngnararathtgyathtgyaaygargarytnaaaygtnaay 591
          || | |  | :|:|  | ||| :|| | |  :| |  | |  :|  | | | |::||  | | ||
Db  83283 GAGGAGAAAATTGAAGCTGGCTGCCAGAAAATTTACATAAGTAATGAGGACCTGAATGTTAAT 83342

Qy    592 cnncargayaaygyngnaraayathwsntggacntgycarmgnwsnwsncararwsnath 651
          | || | |  |: | | | ||:|| | |||  ||:|| | | ||||| | || |||  ||
Db  83343 CCCCAAGACAATGGGAGAAAATGTCTCCAAAGCCATGTCAGAAGGTCTTAACACAGCCCCTA 83402

Qy    652 aarwsnytngcntggmgnccnmgnmgnaaartggttytgyggnaacnggncnggnwsnytn 711
          | |:||:|| |:| | :| ::| :| |  |||:| || |   ||| ||| |  ::|  |
Db  83403 CCATCACAGGCCTGAAGGCCTGAGAGAAAATTGGTTTTCGGGCCGCCCCCAGGGTCCCTG 83462

Qy    712 tgytgygytncarccnnmgnayytngtncnccntgygytnccngtnaaays--------- 758
          ||:||:|    | || :| | ||:||:| ::|  | ||:|| ||| |||:
Db  83463 TGCTGTGTGCAGTCTAGGTACTTTGGTGTCCCTGCATCCTTAGCCCACTCCAGACATGACTAAA 83522

Qy    759 -------ngcngttngcnwsngargggngcnwsnccnaaarccntggcar 798
                  |:| |: |:|| | || | : ||| | | | ||: | || |
Db  83523 AGGGGCCAAGGTACAGCCTCAGGCTGTGTTGCTTCAGAGGGTTGGAAGCTGCAAGCCTTGGCGG 83582

Qy    799 ytnccnwsnggngtngarccncngtnggncnaaraarwsnmgnaathgargtntggarccn 858
          :||| | |  | | |: || |  || | || | | ||:||:|||| |:| ||| ::| | ||
Db  83583 CTTCAAAATGGTATTGAACCTCGCGGGGTTCACAGAAGTCAAGAATTAAGGTTTGGGAACCT 83642

Qy    859 ccnathmgnttycaararathtaygnaaayccntgatgcnmgncaaraarttygngtn 918
          || || :|:|| | |  ||:|:|| |:|  | || |||  |  : | | ||| ||:||
Db  83643 CCGCCCTAGATTTCAGAAAATGTATGGAAAATGCCTGGATGCCAGCGAGAATTTTGCTGCA 83702

Qy    919 gggtnggnwsnsntggmgnacnwsncngmngtngtncaraargnaaygtnggntgg 978
          ||| |  |  |: |:|| :| || | |  || | || | || ||| ||: ||  ||
Db  83703 GGAGTGGGGCCCTCATGGAGAACCTCTGCTAGGGACGTGAGAGCGCGAAAGGGAAAGTGAGGTCA 83762

Qy    979 garccnccncaymgngtnccnwsngggcncccnwsnwsnmgncngtnmgmgnwsnccn 1038
          || |  || | :|:|| ||| |  || | | ::| ||:||:|| ||| :|  ||||
Db  83763 GAGCCCCCAACAGAGCCCCTACTGGGGCACCACCTAGTGGAGCTGTGAAAAGAGGGACA 83822

Qy   1039 ccnwsnwsnmgnytncararggnmgnwsnacngawswsnytncaracaygtnccngaraar 1098
          || || || :|:|||| |  ||  ||||| | |  | | :||:||| | || || || | |
Db  83823 CTGTCCTCCAGACCCCAGAATCGTAGAATCCACTGAGAGCTTGCCACCATGCCACTTGGAAAA 83882

Qy   1099 wsnacngayacncaartgycarccngtnaaargcngcngngnatggaarwsngtnccntayaar 1158
          || | |||| || | |: | | ||| |  ||:||:|| ::|||||| ||:|| ||:| ||||
Db  83883 GCCACAGACACTGAACACACCAGAACCATGAAAACCAGGCAGGGGTGGCCTGTACCCTACAAA 83942

Qy   1159 acngtngtngcngarytnacnaaracngnggnaathtayytnytncaytgycaygayytn 1218
          || ||:||:|| || :||| | | ||:|:|| || |||:|| |||||| |  | |::||
Db  83943 GCCACAGGGTGAAGTTGGCCAAGACCACAGGAACCCACCTCTTGCATCAGCATGACTTG 84002

Qy   1219 gaygtnmgncayggngntnaarmgnayacayttyggngncytnmgcnttyggaytgycnaacn 1278
          || ||: |:||:||:|:||| | :| | |  |||| || | :|| |||||| || | | || |
Db  84003 GATGTGAGACATGGTGTGTCAAAGACACATTATTTTTGGACTTTAAGATTTGACTTGCCACT 84062

Qy   1279 ggnttymgnacntayatggngnccngt-nccnytntgytyggncaarttytcycnttygg 1337
          ||:||: |:||:|| |:||| |::|| |   ||:|:||:|| |  ||:||:| || |||
Db  84063 GGATTTTGGACTTGCACGGGCGCTGTAAACCCCTTGTTTGGGCCAATGTCTCTCATTTGG 84122

Qy   1338 nacngcngtnttyacncaartgyytnttayytncaytgyat 1376
          || || ||:|| | |  ||:|: :|  | ::|  |:|: |
```

```
Db   84123  ATTTGCTGTATTTATCCAATGCCTGTACTGCCAATGTTT 84161

RESULT   8
ACCESSION    AC025233/c
LOCUS        AC025233
DEFINITION   Homo sapiens chromosome 17 clone RP11-333E1, WORKING DRAFT
             SEQUENCE, 27 unordered pieces.
ACCESSION    AC025233
VERSION      AC025233.4  GI:8954317
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 161575)
  AUTHORS    Waterston,R.H.
  TITLE      The sequence of Homo sapiens clone
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 161575)
  AUTHORS    Waterston,R.H.
  TITLE      Direct Submission
  JOURNAL    Submitted (07-MAR-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      On Jul 7, 2000 this sequence version replaced gi:7801489.
             --------------- Genome Center ---------------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site: http://genome.wustl.edu/gsc/index.shtml
             Contact: Project Information ------------
             Center project name: H_NH0333E01
             ----------- Summary Statistics -----------
             Sequencing vector: M13; plasmid: 10%
             Sequencing vector: plasmid; 10%
             Chemistry: Dye-primer ET; 81% of reads
             Chemistry: Dye-terminator Big Dye; 19% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 14558 bases at least Q40
             Consensus quality: 149743 bases at least Q30
             Consensus quality: 152625 bases at least Q20
             Insert size: 176000; agarose-fp
             Insert size: 158975; sum-of-contigs
             Quality coverage: 3.48 in Q20 bases; agarose-fp
             Quality coverage: 3.75 in Q20 bases; sum-of-contigs
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 27 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *
             *      1 1141: contig of 1141 bp in length
             *   1142 1241: gap of unknown length
             *   1242 2524: contig of 1183 bp in length
             *   2425 2524: gap of unknown length
             *   2525 3961: contig of 1437 bp in length
             *   3962 4061: gap of unknown length
             *   4062 5712: contig of 1651 bp in length
             *   5713 5812: gap of unknown length
             *   5813 8019: contig of 2207 bp in length
             *   8020 8119: gap of unknown length
             *   8120 10208: contig of 2089 bp in length
             *  10209 10308: gap of unknown length
             *  10309 11734: contig of 1426 bp in length
             *  11735 11834: gap of unknown length
             *  11835 13265: contig of 1431 bp in length
             *  13266 13365: gap of unknown length
             *  13366 14425: contig of 1060 bp in length
             *  14426 14525: gap of unknown length
```

```
             * 14526 15609: contig of 1084 bp in length
             * 15610 15709: gap of unknown length
             * 15710 17529: contig of 1820 bp in length
             * 17530 17629: gap of unknown length
             * 17630 20351: contig of 2722 bp in length
             * 20352 20451: gap of unknown length
             * 20452 23753: contig of 3302 bp in length
             * 23754 23853: gap of unknown length
             * 23854 26537: contig of 2684 bp in length
             * 26538 26637: gap of unknown length
             * 26638 31339: contig of 4702 bp in length
             * 31340 31439: gap of unknown length
             * 31440 37041: contig of 5602 bp in length
             * 37042 37141: gap of unknown length
             * 37142 45268: contig of 8127 bp in length
             * 45269 45368: gap of unknown length
             * 45369 52323: contig of 6955 bp in length
             * 52324 52423: gap of unknown length
             * 52424 60239: contig of 7816 bp in length
             * 60240 60339: gap of unknown length
             * 60340 68385: contig of 8046 bp in length
             * 68386 68485: gap of unknown length
             * 68486 78834: contig of 10349 bp in length
             * 78835 78934: gap of unknown length
             * 78935 90422: contig of 11488 bp in length
             * 90423 90522: gap of unknown length
             * 90523 100763: contig of 10241 bp in length
             * 100764 100863: gap of unknown length
             * 100864 111672: contig of 10809 bp in length
             * 111673 111772: gap of unknown length
             * 111773 126855: contig of 15083 bp in length
             * 126856 126955: gap of unknown length
             * 126956 142557: contig of 15602 bp in length
             * 142558 142657: gap of unknown length
             * 142658 161575: contig of 18918 bp in length
FEATURES             Location/Qualifiers
     source          1..161575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-333E1"
BASE COUNT    44288 a   35882 c   35114 g   43686 t    2605 others
ORIGIN
```

```
Query Match     35.5%;    Score 488.8;   DB 2;   Length 161575;
Best Local Similarity   48.5%;   Pred. No. 1.5e-106;
Matches  485;  Conservative  202;  Mismatches  280;  Indels  34;  Gaps  2;

Qy      390 rgrgarathwsnaarcarwsnathcargargtnacntgggtnytnytnaargcnt 449
            |:::||:||||:||::||:|:||| |:||:|:||:||:|||:|:|||:||:||||
Db   121612 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGGGTCGTCTTGTTAAATGGATT 121553

Qy      450 ywsnttyathmngargcngatcayaarwsnwsngaraayytncayccngayaaygtnat 509
            |||:|:|| ||:|:||:|||:||:||:||:|| |||||||:|||||||:|||||||||
Db   121552 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTGCAGCCTGAGAAATGTGAT 121493

Qy      510 haaraaraarayccnttywsngargnaaartyaarytngcngcngarathtgyathg 569
            |:::| |::|||:::| |:::||| ||:||:|::|::|:|||||:||:|||:||:|
Db   121492 AGAAAAGAAAAACCCATTTTTCTGAGGAGAAAATTCAATCAGGCTGCAGAAAATTTGCATAAG 121433

Qy      570 yaaygargarytnaaayccncargayaayggngaraayathwsntggacntgyca 629
            |||:|:||:||:|||||||:||:|||||||:| |||||:||:|||||:|||||||
Db   121432 TAACAAGAAGCTGAATGTTAATCCGCAAGACAAATAGGGAAAAATGTCTCCAGGGCATGTCA 121373

Qy      630 rcngnwswsncarccarwsnathaarwsnytngcntggncccnmgngngnaatcggtytg 689
            |||||:|:::||:||||:||||::|||:| |:|:|||||||||:| :|||||:|||||||
Db   121372 GAAGTCTTCACACGCAGCCCCTCCATCACAGGCCTGGAGCCTTAGGAGAGAAAATGGTTTTG 121313

Qy      690 yggnacnggnccnggnwsnytntcgytgygtncarccnmgngayytngcgytncc 749
            |||:|:||||::|||||:|::||::|:::|||:|:|:|||||||:||||:|::||
Db   121312 TGGGCCCAGGCCCCAGGTCCCCAATGCTGTGTCCAGCCCTAGGGACTTGGTTCGCGGTGCATCCC 121253
```

```
Qy    750 ngtnaaywsngcngtngnwsnaarg-----------------------g 776
          : |   | |   ||  | |                               :|
Db 121252 AGCCGCTCTAGCCATGGCTGAAAGGGCCAAATGTAGATAGAGCTGCAAGCTGTGGCTTCAG 121193

Qy    777 ngcnwsnccnaarccnctggcarytnccnwsnggngtngarccngtnggngcnaaraarws 836
          | ::| |||||||:|||  ||:| |||:|||:|||||| |||  | :  ||||:::
Db 121192 AGGGTGCCTAAGCCTACCAGTTTCCACATGGTGTTGTTGAGTCTTTTAGGTACACAGAAGTC 121133

Qy    837 nmgnathgargtntgggarccnccnnathngnttycaararathtaygngnaayccntggat 896
          | ||:||:|| ||||||||||  | |  ||||||| ||| |  | ||| || ||||||||
Db 121132 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAAACGCTTGGAT 121073

Qy    897 gccnmgncaaraaattygcngtngngtnggngtngnwsnstggmgnacnwsngcnmngngt 956
          |  :| ||||||:||||||||  | ||:||| ||:|||  ||:| |||||| | |||
Db 121072 GCCCAGCCAGAAGTTTGCTGCCAGGGCAGGGATCATGGAGAACCTCTCCTAGGGCACT 121013

Qy    957 ncaraargngnaaygtngngtnccnwsngngtnggngncnccnwsnws 1016
          |:||  ||| | | ||||||||| || |||  ||  || : :||
Db 121012 GCAGAAAGGGAAAATGTGGGGTGCGAAGGCCACCATATAGGAGTCCTACTCGGGCCACTTCTAG 120953

Qy   1017 nmgngncngtnmgnmgnwsncnccnwsnwsmgnytncaraargngmgnwsnacngayws 1076
          | |   | |  || | | :|  | |: | |||| |:|:||||| | : :|
Db 120952 TGGGACTGTGAGAAGAGGGCCAATGACCTCCAGAAATGGTAGATCCCGAGAATGGTAGATCCAC 120893

Qy   1077 nytncaryqtnccngaraarwsnacngayacncartgycarccngtnaargngncngg 1136
          | :||  ||| |||  ||||:||:|||  | ::| |:|  |||| || :|
Db 120892 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 120833

Qy   1137 natggarwsngtncntayaaracngtngtngcngarytnacncaaracngtnggngnathta 1196
          |||| |||| |:|| | :| ||||||:|||||| | |||| | |||:|||| ||||||:||
Db 120832 GAGGGAGCTGTGTCCCTACAAAAGCCACACAGGGGTGGAGTGTGCCAAGACCATGGAACCCA 120773

Qy   1197 yytnytncaytgycaygaytngaytngmncayggngtnaarmgngaycaytgyggngc 1256
          |  | |  | |:| | || :|:| |  | |||| |  |||| | |:||| |||||:|
Db 120772 CCTCTTGCATCAGCATGGAGTGTGAGGAGATCATTTGGAGG 120713

Qy   1316 tyggncattgytyccnttyggnacngcngtgtnttyacncar 1356
          |  |||:|| :|:: | |  | ||| | || ||||| ||||:
Db 120712 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 120653

Qy   1257 nytnmgnttygaytgyccnacngngnttymgnantayatgggnccngt-nccnytntgyt 1315
          | :|  | |:|:| ||  ||| :||||| ::||  || |||| |||| |||||:|:|
Db 120652 TTGGCCAAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 120612
```

REFERENCE 2 (bases 1 to 163674)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
           Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
           Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
           Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
           Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
           Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
           Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
           McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
           Murphy,L., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
           O'Connor,T., Naylor,P., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
           Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
           Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
           Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
           Soughez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
           Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
           Zembek,L., Zimmer,A. and Zody,M.

Qy   390 rgargarathwsnaarcacarwsnathcaargargtnacntgggtnytnaargncntt 449
           :|:|||:|||::|| | ||| :| ||||| ::| | |||||||||||:|||:|: | ||

```
           *         1    3134: contig of  3134 bp in length
           *      3135    3234: gap of      100 bp
           *      3235   27783: contig of  24549 bp in length
           *     27784   27883: gap of      100 bp
           *     27884   50218: contig of  22335 bp in length
           *     50219   50318: gap of      100 bp
           *     50319   54526: contig of   4208 bp in length
           *     54527   54626: gap of      100 bp
           *     54627   60200: contig of   5574 bp in length
           *     60201   60300: gap of      100 bp
           *     60301   70374: contig of  10074 bp in length
           *     70375   70474: gap of      100 bp
           *     70475   79596: contig of   9122 bp in length
           *     79597   79696: gap of      100 bp
           *     79697   86804: contig of   7108 bp in length
           *     86805   86904: gap of      100 bp
           *     86905  123464: contig of  36560 bp in length
           *    123465  123564: gap of      100 bp
           *    123565  160329: contig of  36765 bp in length
           *    160330  160429: gap of      100 bp
           *    160430  163674: contig of   3245 bp in length
```

```
FEATURES             Location/Qualifiers
     source          1..163674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-333E1"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /map="17"
BASE COUNT     44782 a    37406 c    35496 g    44882 t    1108 others
ORIGIN
```

Qy   390 rgargarathwsnaarcacarwsnathcaargargtnacntgggtnytnaargncntt 449
           :|:|||:|||::|| | ||| :| ||||| ::| | |||||||||||:|||:|: | ||

Query Match: 35.5%; Score 488.8; DB 2; Length 163674;
Best Local Similarity   48.5%; Pred. No. 1.5e-106;
Matches 485; Conservative 202; Mismatches 280; Indels 34; Gaps 2;

```
RESULT 10

Db 136131 GGAAGAAAATTTCTPAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCGTCCTGTTAAATGGATT 136190
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    450 ywsnttyathmgnargcngarcayaarwsnvsngaraaaygtnacaycongayayaaygtnat   509
Db 136191 CAGTTTTATGAGGGAAGCAGAGCATAAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 136250

Qy    510 haararaaraaycconttywsngarggnaaarttyaarytngcngncngaaathcgyaththg   569
          : :: ||||||:|| |||||: || ||||||:|| |||:|| ||||||||||||:||:||:
Db 136251 AGAAAAAGAAAAACCCATTTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAAATTTGCATAAG 136310

Qy    570 yaaygargaarytnaaygtnaayccncargayagayaaayathwsntggacntgyca   629
          :||| || ||| ||||: || || ||||| || |||| || :|| |||||:|| ||:
Db 136311 TAACAAGAGACTGAAATGTTAAATCCGAAGACAATAYGGGGAAAAATGTTCTCCAGGGCATGTCA 136370

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaaartggttytg   689
          :|| || ||:||:|| |||||:||| || ||||||:||||:||:||:|| |||||||||:
Db 136371 GAAGTCTTCACACAGCACCCCTCCCATCACAGGCCTGGAGGAGAAAAATGGTTTTG 136430

Qy    690 yggnacnggncanggnwsnytntgytgygtncarccnmngayytngtncentgygtncc   749
          :||||| |||||| || ||:|:|:|:|: |||||||:|| |:||:|||||:|||: ||
Db 136431 TGGGCCAGGCCCAGGGTCCCAAATGCTGTGTCCAGCCTAGGGACTTGGTCGCGTGCCATCCC 136490

Qy    750 ngtnaaaywsngcngtngcnwsngarg------------g   776
          |:|||| :|| || ||:|| || ||||            :|
Db 136491 AGCCGCTTCTAGCCATGGGCTGAAAAGGGCCCAATGTAGAATAGAGCTTCAAGCTGTGGCTTCAG 136550

Qy    777 ngcnwsnccnaarccntgggcaryytnccnwsngngntgnarccngtngngcnaaraarws   836
          |||| |||||||:|||||||||: ||||||||:|:|:|||||||||:|:|||||||||:
Db 136551 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 136610

Qy    837 nmgnathgargtntgggarccncncnaathmgntycaraaaratttaygggnaaycontggat   896
          |:||||:||||:|||||||||:||:||||||:|:||||||:| |||||||||||:||:|||||
Db 136611 AAGAATTCAGGTTTGGGAACTTCCACCTGGATTTCAGAGAGATGTTATGGGAAAACGCTTTGGGAT 136670

Qy    897 gccnmgncaraarttygcngtnggngtnggngnwsnsntggmgnacnwsngcnmgngtngt   956
          |||:|:|||:|||||:||:|||||:|||||:|||| |:|||:|||||| |||:|:|:|:|:
Db 136671 GCCCAGCCAGGAGGGAAATGTTGCTACAGGGCAGGGATCTCATGGAGAACCTCTTCCTAGGCACT 136730

Qy    957 ncaraaaargnaaaytnggngntgggngtncenwsnggngcnccnwsnws  1016
          ||:||||||:||||:||||:|:|||||:|:|||:|||||||:||:||:||
Db 136731 GCAGAAGAGGGAAATGTGGGGTCGGAAGCCACCATATGGAGTCCCTACTCGGCGCACCTTCTAG 136790

Qy   1017 nmgngcngctnmgnamgnwsnccnccnwsnwsmngnytncaraaargngmgnwsnacngayws  1076
          |:|:|| |||:|:|:|:|| ||:||:|| ||:||:|:||||||||:|:|:|||| ||:|||||
Db 136791 TGGGAGCTGTGAGAAGAGGGCCCAATGACCTCCAGACCCCAGAATGGTGTAGAGTCACCTACAG 136850

Qy   1077 nytncarcaygtncngaraarwsnacncngayacncaartgycaarccngtnaargcngcngg  1136
          |:|||||||:|:||:|||:||||:||:|:||:||:||||:|:||||||:|:|||:|:|:|:||
Db 136851 CTTGCACCGTATAGCCTGGAAAAAGCTCAGACACTCAACACCCGTGAAAGCACCAG 136910

Qy   1137 natggarwsngtncontayaaracngtngtngcngaryatnacnaacacngtnggnathta  1196
          ||:|||| |:|:|:||:|||||:|:|:|:|:|:||||:|:|||:|||||:|:||:|||:||
Db 136911 GAGGGAGGCGTGCCCTACAAAAGCCAGAGGTGGGAGTTGCCCAAGACCATGGGAACCCA 136970

Qy   1197 yytnytncaytgycaygayytngaytnggngcaytnaaarmgngaycaytyggngnc  1256
          :|:|:||||:|:||:|:||:|||||:|||||||:|:||||:|:|||||:||:|:|:|
Db 136971 CCTCTTGCACTGGAGTGACCTGGAYGTGACAACCTGGGAYCAGANGATCATTTTGGAGG 137030

Qy   1257 nytnmgnttygaytgyccnacncngngnttymgnacntayatggngnccngt-ncnytntgyt  1315
          |:|:|:|:||||:|:||||:|:|:|:|:||:|:|||:|||:|||:|:||:|:: |:|:|:|:|:
Db 137031 TTTAAAAATTTGATTNCGTTCGTGGATTTCGGACTTGCATGGGCCTGTACCCCTTTGTT 137090

Qy   1316 tyggncaarttyccnttyyccntyggnacngcngtnttyacncar  1356
          |:||:|||||||||:||::|||:|:|||||:|:|:||:||||:||:
Db 137091 TTGGCCAAATTTCTCCCAATTTGGAACAGCTGTATTTACCCAG 137131
```

```
FEATURES             Location/Qualifiers
     source          1..175639
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-420A6"
     misc_feature    1..13947
                     /note="assembly_fragment"
                     clone_side:SP6
     misc_feature    14048..14259
                     /note="assembly_fragment"
                     vector_side:left"
     misc_feature    14360..15039
                     /note="assembly_fragment"
     misc_feature    15140..15832
                     /note="assembly_fragment"
     misc_feature    15933..16689
                     /note="assembly_fragment"
     misc_feature    16790..18092
                     /note="assembly_fragment"
     misc_feature    18193..19993
                     /note="assembly_fragment"
     misc_feature    20094..22205
                     /note="assembly_fragment"

*    19994 20093: gap of 100 bp in length
*    20094 22205: contig of 2112 bp in length
*    22206 22305: gap of 100 bp in length
*    22306 23915: contig of 1610 bp in length
*    23916 24015: gap of 100 bp in length
*    24016 26544: contig of 2529 bp in length
*    26545 26644: gap of 100 bp in length
*    26645 28851: contig of 2207 bp in length
*    28852 28951: gap of 100 bp in length
*    28952 30454: contig of 1503 bp in length
*    30455 30554: gap of 100 bp in length
*    30555 33015: contig of 2461 bp in length
*    33016 33115: gap of 100 bp in length
*    33116 35727: contig of 2612 bp in length
*    35728 35827: gap of 100 bp in length
*    35828 59255: contig of 23428 bp in length
*    59256 59355: gap of 100 bp in length
*    59356 62539: contig of 3184 bp in length
*    62540 62639: gap of 100 bp in length
*    62640 67930: contig of 5291 bp in length
*    67931 68030: gap of 100 bp in length
*    68031 73174: contig of 5144 bp in length
*    73175 73274: gap of 100 bp in length
*    73275 77427: contig of 4153 bp in length
*    77428 77527: gap of 100 bp in length
*    77528 81788: contig of 4261 bp in length
*    81789 81888: gap of 100 bp in length
*    81889 86379: contig of 4491 bp in length
*    86380 86479: gap of 100 bp in length
*    86480 91547: contig of 5068 bp in length
*    91548 91647: gap of 100 bp in length
*    91648 97783: contig of 6136 bp in length
*    97784 97883: gap of 100 bp in length
*    97884 105190: contig of 7307 bp in length
*    105191 105290: gap of 100 bp in length
*    105291 111855: contig of 6565 bp in length
*    111856 111955: gap of 100 bp in length
*    111956 119820: contig of 7865 bp in length
*    119821 119920: gap of 100 bp in length
*    119921 131299: contig of 11379 bp in length
*    131300 131399: gap of 100 bp in length
*    131400 139546: contig of 8147 bp in length
*    139547 139646: gap of 100 bp in length
*    139647 155538: contig of 15892 bp in length
*    155539 155638: gap of 100 bp in length
*    155639 173481: contig of 17843 bp in length
*    173482 173581: gap of 100 bp in length
*    173582 175639: contig of 2058 bp in length.

     misc_feature    22306..23915
                     /note="assembly_fragment"
     misc_feature    24016..26544
                     /note="assembly_fragment"
     misc_feature    26645..28851
                     /note="assembly_fragment"
     misc_feature    28952..30454
                     /note="assembly_fragment"
     misc_feature    30555..33015
                     /note="assembly_fragment"
     misc_feature    33116..35727
                     /note="assembly_fragment"
     misc_feature    35828..59255
                     /note="assembly_fragment"
     misc_feature    59356..62539
                     /note="assembly_fragment"
     misc_feature    62640..67930
                     /note="assembly_fragment"
     misc_feature    68031..73174
                     /note="assembly_fragment"
     misc_feature    73275..77427
                     /note="assembly_fragment"
     misc_feature    77528..81788
                     /note="assembly_fragment"
     misc_feature    81889..86379
                     /note="assembly_fragment"
     misc_feature    86480..91547
                     /note="assembly_fragment"
     misc_feature    91648..97783
                     /note="assembly_fragment"
     misc_feature    97884..105190
                     /note="assembly_fragment"
     misc_feature    105291..111855
                     /note="assembly_fragment"
     misc_feature    111956..119820
                     /note="assembly_fragment"
     misc_feature    119921..131299
                     /note="assembly_fragment"
     misc_feature    131400..139546
                     /note="assembly_fragment"
     misc_feature    139647..155538
                     /note="assembly_fragment"
     misc_feature    155639..173481
                     /note="assembly_fragment"
     misc_feature    173582..175639
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:right"

BASE COUNT   48575 a   37483 c   38030 g   48544 t   3007 others

Query Match    35.5%;  Score 488.8;  DB 2;  Length 175639;
Best Local Similarity   48.5%;  Pred. No. 1.5e-106;
Matches  485;  Conservative  202;  Mismatches  280;  Indels  34;  Gaps  2;

Qy        390 rgargarathwsnaarcaarcarwsnathcaargargtnacntgggtnytnaytnaargcntt 449
              :!!!:!!::!!:::: !:!!:!:!    !!!!!!!! :: !:!!:!:  ! :!! !! ! !!
Db       2929 GGAAGAAATTTCTTAAGCACAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 2870

Qy        450 ywsnttyathmgngargcngaycayarwsnwsngaraayytncayccngayaaaygtnat 509
              ::: :!!!:!!:!! :!:!!:!!:!::!::: !!!!!!!!!!!!!!! !!:! !!:!!!!
Db       2869 CAGTTTTATGAGGGAACGAGAGCCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 2810

Qy        510 haararacaaraaayccnttywsngargngnaarttyaarytngcngararathtgyathtg 569
              :!!:!!!:!!:! !!! :! :!!!!!!:!!!!:!!:! !!! !!::!!!!:!!:!
Db       2809 AGAAAAAGAAAAACCCATTTTCTGAGGAGAGAAAATTCAATCAGGCTGCAGAAAATTTGCATAAG 2750

Qy        570 yaaygargaarytnaaygtnaaycccnacargayaaaygggnaraayathwsntggacntgyca 629
              !!!: !:! :!!:! !:!!! !!!:!!: !!!!!!!!!:!!:!:!!:!   !!:! !!!!!
Db       2749 TAACAAGAAAGCTGAATGTTAATCCACAGACAATGGGGAAAAATGTCTCCAGGCATGTCA 2690

Qy        630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
              :: ::!! !:!:!:!!: !: .::: ! :!!:!!!!!!!!:!!!!!!!:!!!!!!!!!
Db       2750                                                            689
```

```
Db  2689  GAAGTCTTCACAGAGCCCTCCATCACAGGCCTGGAGGCCTAGGAGGAAAATGGTTTTG 2630
Qy   690  yggnacnggngncnggnwsnytntgytgyytncarcacnmgnagayytngtnccntgygtncc 749
Db  2629  TGGGCCAGGCCCAGGGCTCCCAATGCTGTGTGTTCCAGCCTAGGGACTTGGTGCGCTGCATCC 2570
Qy   750  ngtnaaywsngcngtngcnwsngarg---------------------------g 776
Db  2569  AGCCGCTCTAGCACCATGGCTGAAAGGGCCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 2510
Qy   777  ngcnwsnccnaarccntggcarytncccnwsnggngntnggncgncnaaraarws 836
Db  2509  AGGGTGCCCTAAGCCCTTGGCAGTTTCCACAATGGTGTTGAGTGTTTAGGTACACAGAAGTC 2450
Qy   837  nmgnathgargtntggygarccncncnathmgnttycaraaraathtaygnaaycntggat 896
Db  2449  AAGAATTGAGGTTTGGGAACTTCACCTGGATTTCAGAACAGTGTAInGGAAACGCTTGGAT 2390
Qy   897  gccnnmgncaraarttygcngtnggntnggnnwsnwsntggmgnacnwsngcnmgnmgtngt 956
Db  2389  GCCCAGCCAGAAGTTTGTCGCAGGGGCAGGGGATTCATGGAGAACCTCGCTAGGGCACT 2330
Qy   957  ncaraaaargnaaygtnggntgggatccncncnacayngngtnccnwsnggngcncnwsnws 1016
Db  2329  GCAGAAGGGAAAATGTGGGGTGCGAAGCACACCATATGGAGTCCCTACTCGGGCACTTCTAG 2270
Qy  1017  nmgngncngtnmgnmngnwsnccncncnwsnwsnmgnytncaraarggnmgnwsnacngayws 1076
Db  2269  TGGACGTGTGAAGAAGGGGCCCAAATGACCTCCAGACCCCAGAATGGTGGTAGAGTCCAACTACAG 2210
Qy  1077  nytncarcaygtncngaraarwsnacnacngayacnnaarwsntaarcgncngcngg 1136
Db  2209  CTTGCACCGPATGCCTGGAAAAAGCTGCAGACACTCAAACACACCAGCCCGTGAAAAGCAGCCAG 2150
Qy  1137  natggarwsngtncnctayaaracngtngtngcngarytnacnaaracngtnggnathta 1196
Db  2149  GAGGGAGGCGTGCCCTACAAAGCCAGGGGTGGAGTTGCCCAACCATGGAACCCA 2090
Qy  1197  yytnytncaytgycaygayytngaygtnmcnaygngtnaarmgngayayttyyggngc 1256
Db  2089  CCTCTTCCATCAGCATGACCTGGATGTGAGCTCAAAGGAGATCATTTTGGAGG 2030
Qy  1257  nytnmgnttygaytgyccnacngnnttymgnacntayatggggncngt-nccnytntgyt 1315
Db  2029  TTTTAAAATTTGATTGCTGTGGATTTCGGACTTGCCATGGGCCCTGTAACCCCTTTGTT 1970
Qy  1316  tyggncartyttyycnttyygnacngncngtnttyacncar 1356
Db  1969  TTGGCCAATTTCTCCCATTGGAACAGCTGTPATTTACCCAG 1929
```

RESULT 11
AL391823/c

```
LOCUS       AL391823    181842 bp    DNA    linear    HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-319F11, *** SEQUENCING IN
            PROGRESS ***, 12 unordered pieces.
ACCESSION   AL391823
VERSION     AL391823.9  GI:11995113
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181842)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 28, 2000 this sequence version replaced gi:11990058.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
```

Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-------------- Project Information
Center project name: bA319F11
-------------- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 95% of reads
Chemistry: Dye-terminator ET-amersham; 3% of reads Chemistry:
Dye-terminator Big Dye; 96% of reads
Consensus quality: 174456 bases at least Q40
Consensus quality: 177550 bases at least Q30
Consensus quality: 179392 bases at least Q20
Insert size: 180742; sum-of-contigs
Insert size: 162786; 7.7% error; agarose-fp
Quality coverage: 4.40x in Q20 bases; sum-of-contigs Quality
coverage: 4.89x in Q20 bases; agarose-fp
----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*              1  8896: contig of 8896 bp in length
*           8897  8996: gap of       100 bp
*           8997  22635: contig of 13639 bp in length
*          22636  22735: gap of       100 bp
*          22736  55625: contig of 32890 bp in length
*          55626  55725: gap of       100 bp
*          55726  65440: contig of 9715 bp in length
*          65441  65540: gap of       100 bp
*          65541  88588: contig of 23048 bp in length
*          88589  88688: gap of       100 bp
*          88689  94602: contig of 5914 bp in length
*          94603  94702: gap of       100 bp
*          94703  105410: contig of 10708 bp in length
*         105411 105510: gap of       100 bp
*         105511 107644: contig of 2134 bp in length
*         107645 107744: gap of       100 bp
*         107745 126586: contig of 18842 bp in length
*         126587 126686: gap of       100 bp
*         126687 155610: contig of 28924 bp in length
*         155611 155710: gap of       100 bp
*         155711 173103: contig of 17393 bp in length
*         173104 173203: gap of       100 bp
*         173204 181842: contig of 8639 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..181842
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-319F11"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..8896
                     /note="assembly_fragment:02130"
                     /clone_end:SP6
                     vector_side:left"
     misc_feature    8997..22635
                     /note="assembly_fragment:01909
                     fragment_chain:1"
     misc_feature    22736..55625
                     /note="assembly_fragment:00961
                     fragment_chain:1"
     misc_feature    55726..65440
                     /note="assembly_fragment:00129
                     fragment_chain:1"
     misc_feature    65541..88588
                     /note="assembly_fragment:00944
                     fragment_chain:1"
     misc_feature    88689..94602
```

```
misc_feature    94703..105410
                /note="assembly_fragment:01651"
misc_feature    105511..107644
                /note="assembly_chain:1"
misc_feature    107745..126586
                /note="assembly_fragment:01169"
misc_feature    126687..155610
                /note="assembly_chain:1"
misc_feature    155711..173103
                /note="assembly_fragment:01445"
misc_feature    173204..181842
                /note="assembly_chain:1"

BASE COUNT      53809 a    39941 c    37882 g    49105 t    1105 others
ORIGIN
```

```
Query Match       35.5%;   Score 488.6;   DB 2;   Length 181842;
Best Local Similarity    47.5%;   Pred. No. 1.7e-106;
Matches    486;   Conservative 212;   Mismatches 290;   Indels 36;   Gaps 2;

Qy    389  argargarathwsnaacarcarwsnathcargargtnaacntgggtnytnytnaaargcnt  448
Db  64835  AAGAAGGAGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTGTGTTAAGGCAT  64776
Qy    449  tywsnttyathmngargcngarcaycayaaarwsnwsngaraayytncaycayayaaygtna  508
Db  64775  TCAGTTTTTATAAGGGAAGCAGATCATAAAAAATTTGGAAAATTTGCACCTGACTATGCAA  64716
Qy    509  thaararaaraayccnttywsngargnaarttyaarytngcngcngararathtgyatht  568
Db  64715  TAGAAAAGAAAAACCATTTTCTGGGGAGAAAATTCAAGCCTATTGCAGAAAATTTGCCTAA  64656
Qy    569  gyaaaygargargarytnaaaycncaargayaaaygngaraaayathwsntggacntgyc  628
Db  64655  GTGACAAGGAGCCTAAATGTTAAATCCCCAAGACCATCGGGGAAAATGTCTACAGGCCATGT  64596
Qy    629  armgnwsncarcarwsnathaaarwsnytngcntggmgncncmgmng---naartggt  685
Db  64595  AGGGAACTTCATGGCTCCCCCCTCCCCATCACAAGCCTCGCAGGCGCCAGGAGTAAAAAGTGGT  64536
Qy    686  tytgyggnacnggncnggnwsnytntgytgytgytgtgtncatccnmngnagayytngtnccntgyg  745
Db  64535  TTCATGGCCAGGCCCAGGGGTTCCTGTGCCGTGTCCACAGGGACTTGGTGTCGCCCGTG  64476
Qy    746  tncCng-------------------tnaaywsngcngtngcnwsng  772
Db  64475  TCCCAGCTGCTCCAGCTATGGCTGAAAAGGGCGCAAGGTAGAGGTTGGGCTGTGCCTTCAG  64416
Qy    773  argngncnwsnccnaarccntggcarytncCnwsngggntngarccngtnggngcnaara  832
Db  64415  AGGGTGCAAGAACCCAAGCCTTGGCAGCGTTCCACGTGATGTTCAGCCTGTGGGGTGCACAGA  64356
Qy    833  arwsnmgnaathgargtntgggarccncnathmgnttycaraararathtayggnaaycnt  892
Db  64355  AGTCAAGAACTGAGGTTGGGAAACCTCTGCCTAGATTTCAGAAGATGTATGGAAACGCCT  64296
Qy    893  ggatgcncmgncaraaattygcngtngggntgnggnwsnwsntggmgnacnwsngcnmgng  952
Db  64295  GGATGCCCAGGCAAAAGTTTGCTCGACGTGTGGGACCCTCATGGAGAACCTCTCCTAGG  64236
Qy    953  tngtncaraargnaaygtnggntggarccncncncaymgngtnccnwsngggntcnccnw  1012
Db  64235  CAGTGCAGAAGGAGAAAAATGTGGGGTCAGACGCCCCACCACAGAGTCCCTACTGGGGCCACAC  64176
Qy   1013  snwsnmgnacngtnmgnmgnwsnccnccnwsnsnmgnytncaraarggnnmgnwsnacng  1072
Db   1073  aywsnytncarcaycaygtnccngararwsnacngayacncaartgycarccngtnaargcng  1132
Db  64835  AAGAAGGAGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTGTGTTAAGGCAT  64776

Qy   1073  aywsnytncarcaygtnccngararwsnacngayacncaartgycarccngtnaargcng  1132
Db  64115  ACAGCTTGCACCATTCACCTGGAAAAGCCACAACAGACACTCAACGCGCCATGAAAGCAG  64056
Qy   1133  cngngnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna  1192
Db  64055  CCAGGAGGGGAGGCTGTACCCTGCAAAGCCACACAGGGCGTGGAACTGCCCAAGACTATGGGAA  63996
Qy   1193  thtayytnytcncayctgycaaygayytngaygtnmgncayggngtnaaarmgngaycaytyg  1252
Db  63995  CCTACCTCTTGCATCAGCATGATCAGCCTTGGATCTGAGACCTGGTGTCAAAAGGAGGATCATTTTG  63936
Qy   1253  gngcnytnmgnttygaytgyccnacncnggnttymgnaacntayatgggngncngtnccncnytnt  1312
Db  63935  AAGCTTTAAAATTTGACTGCGTCCTCACTCGGATTTCAGACTTGTCATGGGCCCTGTGTCAAAAGGAGATCATTTTG  63876
Qy   1313  gyttyggncaarttytyccnttyggnaacngcngntyttyacncattgyyntayytncayt  1372
Db  63875  GTTTTTGGGCCAATTTTCTCCCCATTTGGGAGGGGCTATATTTACCCAGTACTTGTACCCCATT  63816
Qy   1373  gyat  1376
Db  63815  GTAT  63812
```

```
RESULT 12
AL353634

LOCUS       AL353634               77405 bp    DNA    linear    HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome X clone RP11-435A2 map q21.33-22.3, ***
            SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AL353634
VERSION     AL353634.4   GI:9863649
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 77405)
  AUTHORS   McIay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9213111.
            ------------------------    Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ------------------------    Project Information
            Center project name: bA435A2
            ------------------------    Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 13% of reads Chemistry:
            Dye-terminator Big Dye; 86% of reads
            Consensus quality: 76604 bases at least Q40
            Consensus quality: 77098 bases at least Q30
            Consensus quality: 77320 bases at least Q20
            Insert size: 77405; sum-of-contigs
            Insert size: 85298; 1.5% error; agarose-fp
            Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
            coverage: 4.30x in Q20 bases; agarose-fp
            ------------------------
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            ------------------------
FEATURES             Location/Qualifiers
     source          1..77405
                     /organism="Homo sapiens"
```

```
misc_feature
     /db_xref="taxon:9606"
     /chromosome="X"
     /map="q21.33-22.3"
     /clone="RP11-435A2"
     /clone_lib="RPCI-11.2"
     1..77405
     /note="assembly_fragment:00554"

BASE COUNT    25072 a    14551 c    15133 g    22649 t
ORIGIN
```

```
Query Match        35.6%;    Score 488.2;    DB 2;    Length 77405;
Best Local Similarity 47.0%;    Pred. No. 1.6e-106;
Matches 508;    Conservative 225;    Mismatches 310;    Indels 37;    Gaps 4;
```

```
Qy    334 ytncaycaytgyaaygtnaayathtggytnaatgmgnaarcararthytgcnaayaargar 393
          :::|||::| :|||||  |||:||||||||||  ::|| | || | |||| ||||||| :||
Db   8637 CTGCCCTAGAGATCGTGTGGGAAVTTGACCTTGAGAGAGATGAAVTTAGGTATCAGGTGGAA 8696

Qy    394 garathwshaaarcaarwsnathcarargargtnacntgggtnyttnytnaargcnttywsn 453
          :||:|| ::||| |||  |||| |||||||||| :|| || |:| ||||||| || |:||::
Db   8697 GAAAVTTTCTAAAGCAAGCAAAAGCATTCAAAGAGGTGACTTGTGTCCTGTGTTAAAAGCACTCAGT 8756

Qy    454 ttyathmgngargcngarcaayaarwsnwsngaraaaytyncaycnyngayaaygtnathaar 513
          ||:  | || ||||| || ||| ||||| || | :||||| |||| |||| |||| |||||:
Db   8757 TTTCCAAGGAAGCAGAACATAAAAGTTCAGAAAAATTTTCAGCCTTGACAAATGTGTGATAGAA 8816

Qy    514 aaraaraaycc-nttywsngarggnaaarttyaarytngcngcngararathtgyathtgyaa 572
          ||| || ||  :|||||||| || ||| ||||| |||:|||| ||| |||| |||||| |||
Db   8817 AAGAAAAACCCATTTTTCTGATGATAAAVTTGAAGCCAGCTGCAGAAVTTTGCATAAGTAA 8876

Qy    573 ygargarytnaaygtnaayccncargayaayggngaaraayathwsntgacntgycaarmg 632
          ||:|||:| :|||| || |  :||||| |||| |||  ||:|||||||| ||||||:|| |
Db   8877 CGAGGAGCCCAAAVTGTTAAVTCACTRAGACCATGGGGAAAAATATCTCCAACGGCTGTCAGAG 8936

Qy    633 nwsnwsncaarcaarwsnathaarwsnytng--cntggmgnccnnmgnaartggtttytgy 690
          : ||:||  ::|| || ||||| | |     :|||||||||||||||||| || |||||:|
Db   8937 GTCTTCACGGCAGCCCCTCACATCACTGGCGCCCCCAGAGGCCTAGGAGAAAATAGGTTTTAT 8996

Qy    691 ggnacngngnccnggnwsnytntgytgyytgncarccnnmgnayytngtnccntgygtnccn 750
          || ||:|| || ||||| :|||:|:||||||| || ||:||||||||| |||||||||||||
Db   8997 GGGTCGTGGGGCCCCAGGGTCCCATGCTATGTGCCAGTCTAGGGACTTGGTCTGCGTGTCCCA 9056

Qy    751 gtnaa------------------------------ywsngcngtngcnwsnarggn 777
          ::|||                              ::|||||||||||||||||||
Db   9057 GCCACTCCAGCCATAACTGAAAGGGGCGGTATAGCTTGGGGCGTTGCCTTCAAAGGAT 9116

Qy    778 gcnwsnccnaarccntggcgaryntccnwsngnggtngarccngtngggngcnaaraarwsn 837
          || | :|| || |||||||||| |||| ||||||||| |||||||| |||||  ::|||||
Db   9117 GGAGCCCCAGGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCCAGGTGCACTGAAGTCA 9176

Qy    838 mgnathgargtnttggarccncncnathmgntycaraaratathtayggnaayccntggatg 897
          :: |||:||||| |||| |  |||  |||:||| || || ||||||||||| ||||:|:|||
Db   9177 AGAAVTTGAGGTTTGGGAACCTCCACTGAGTTCAGAAGAVTGTATGGAAAVTGCCTGGATG 9236

Qy    898 ccnmgncaaraarttygcngtngnggtngggnwsnwsntggngnacwsngcnmgntngtngtn 957
          || ||| ||| | ||||:|||||||||| :||||| ::| | |||| |||| |||||||||:
Db   9237 CCCAGGCAAAAGTTTGCTGCAGGGCGGAGGACCTCATCGGAAGAACCTCGTTAGGTGTGGTG 9296

Qy    958 caraaargnaaygtngggtnggggarccncncaayngngtgyaccncncaywsnccwsnwsn 1017
          ||| |||:|||||||||:||| || ||||||| ||:||:||||  ||||||||||:||:||
Db   9297 TGGAAGGAAAAGTGGGGTCATAGCCGCCACACAGAGTCCCTACCACAGGGGGACCACTAGT 9356

Qy   1018 mgngcngtnmgngmngwsnccncnwsnsnmgnytncaraargnmgnwsnacngaywsn 1077
          :: ||:||| ||:| | || || || |||| ||||| || || ||:|||:|| |||:
Db   9357 GGAGCTGTGAGAGAACAGGGCCCACACATCCTCCAGACAAATGGCCACCGACAGC 9416

Qy   1078 ytncrcaygtnccngararwsnacngayacnccngtngargcngcnggn 1137
          |: |||||||| ||| ||||| || | | ||||| || ||||||||||:
Db   9417 CTGCACCGTGTGCCTGGAAAAGCCCGAGACAACTCAAVTGCCAGCGACCAGG 9476

Qy   1138 atggarwsngtncncntayaaracngtngcngaryacnacngtngnnathtay 1197
          ||||| ||||||| | | |||| ||||:||||| |  | :|||||| :||:||
Db   9477 AGGGAGGCGTGACCCTGCAAAGCCACAGGGGGCGAGAGACTGCCCAAGATCATGGGAACCAAT 9536
```

```
Qy   1198 ytnytncaytgycaygayytngaygtnmgncayggngngtnaarmgngaygaycayttyggngcn 1257
          |:|:||||||||||:|||:||||| | |||| ||| |||||:| ||:||:||:|||:||:| |||
Db   9537 CTCTTGCATCAGCATGACCTGGATATGAGACAVTGGAGGAGTCAAAGGAAAVTCAVTTTGGAGCT 9596
```

```
Qy   1258 ytnmgntttygaytgyccnacnggnttymgnaacntayatgggvnccngt-nccnytntgytt 1316
          |:|  |:|| || |||| || |  |||:|||| |||||||||| ||||: :||||:|||:|||
Db   9597 TTAAGAVTTGACTGTCCCCACTGGGATCTCAGACTTCCAVTGGGGCCGTGTAGGCTTCTTTGTTT 9656
```

```
Qy   1317 yggncarttyycnttygggnacngcngtnattyacncaartgyyytataayytncaytgyat 1376
          ||:||: |||:| |||||:|:|||| ::|||||||||||||:||  ||:|| ||||||| ||
Db   9657 TGGCCAATATCTCCCATTTGGAATGGCTGAVTTATCCAAVTGCCTGTACCCCAVTGTAT 9716
```

```
RESULT 13

HSJ633H17/c

LOCUS       HSJ633H17    124531 bp    DNA    linear    PRI 07-FEB-2000
DEFINITION  Human DNA sequence from clone 633H17 on chromosome 1p31.2-32.2.
            Contains a pseudogene similar to part of MTCO1 (Cytochrome C
            oxidase 1), MTCO2 (cytochrome c oxidase II), MTATP8 (ATP synthase
            8) and GOT2 glutamic-oxaloacetic transaminase 2, mitochondrial
            (aspartate aminotransferase 2) pseudogenes, ESTs and GSSs, complete
            sequence.
ACCESSION   AL049710
VERSION     AL049710.18  GI:5650682
KEYWORDS    HTG; aspartate aminotransferase; ATP synthase; Cytochrome C
            oxidase; glutamic-oxaloacetic transaminase; GOT2; MTATP8; MTCO1;
            MTCO2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124531)
  AUTHORS   Moore,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 29, 1999 this sequence version replaced gi:5565565.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 633H17. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            633H17 is from the library RPCI-4 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 The
            following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.
FEATURES             Location/Qualifiers
     source          1..124531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p31.2-32.2"
                     /clone_lib="RPCI-4"
                     /clone="633H17"
```

```
repeat_region   1. .234
                /note="AluSx repeat: matches 59. .292 of consensus"
misc_feature    complement(235. .546)
                /note="match: GSS: Em:AQ055811"
repeat_region   559. .760
                /note="MIR repeat: matches 74. .262 of consensus"
repeat_region   1158. .2018
                /note="L1MA4 repeat: matches 5282. .6145 of consensus"
repeat_region   2026. .2590
                /note="HERVL repeat: matches 4676. .5241 of consensus"
repeat_region   2879. .9441
                /note="L1PA2 repeat: matches 1. .6146 of consensus"
repeat_region   10000. .10124
                /note="MIR repeat: matches 34. .154 of consensus"
repeat_region   10623. .10689
                /note="MIR repeat: matches 36. .102 of consensus"
repeat_region   11014. .11136
                /note="L2 repeat: matches 2389. .2710 of consensus"
repeat_region   11136. .11262
                /note="L2 repeat: matches 2349. .2473 of consensus"
repeat_region   11340. .13522
                /note="MIR repeat: matches 114. .219 of consensus"
repeat_region   13602. .13725
                /note="L2 repeat: matches 130. .262 of consensus"
repeat_region   13739. .13987
                /note="L2 repeat: matches 2397. .2666 of consensus"
repeat_region   14358. .14486
                /note="L2 repeat: matches 2585. .2705 of consensus"
misc_feature    14556. .14764
                /note="match: GSS: Em:B30406"
repeat_region   14656. .14764
                /note="MIR repeat: matches 124. .252 of consensus"
misc_feature    14765. .14965
                /note="match: GSS: Em:AQ454645"
misc_feature    14966. .15150
                /note="match: GSS: Em:AQ357205"
misc_feature    15157. .15818
                /note="match: GSS: Em:AQ309712"
repeat_region   16403. .16945
                /note="MER39b repeat: matches 14. .579 of consensus"
repeat_region   16410. .16933
                /note="MIR repeat: matches 13. .217 of consensus"
repeat_region   16410. .17037
                /note="23 copies 2 mer tt 74% conserved"
misc_feature    complement(16459. .16861)
                /note="match: GSS: Em:AQ197233"
repeat_region   16531. .16576
                /note="MIR repeat: matches 7. .107 of consensus"
repeat_region   18053. .18154
                /note="MER39b repeat: matches 476. .574 of consensus"
misc_feature    18377. .18492
                /note="match: GSS: Em:AQ210513"
repeat_region   18525. .18954
                /note="MIR repeat: matches 10. .114 of consensus"
repeat_region   19026. .19126
                /note="MIR repeat: matches 28. .169 of consensus"
repeat_region   19866. .20007
                /note="LTR16C repeat: matches 90. .152 of consensus"
repeat_region   20273. .20334
                /note="MER81 repeat: matches 1. .110 of consensus"
repeat_region   20403. .20505
                /note="LTR16C repeat: matches 190. .376 of consensus"
repeat_region   20508. .20690
                /note="18 copies 2 mer ac 94% conserved"
repeat_region   20692. .20725
                /note="L2 repeat: matches 2660. .2693 of consensus"
repeat_region   20692. .20863
                /note="MER5A repeat: matches 40. .167 of consensus"
repeat_region   21172. .21304
                /note="MER5A repeat: matches 56. .188 of consensus"
repeat_region   21201. .21331
                /note="MER54A repeat: matches 214. .285 of consensus"
repeat_region   21466. .21528
repeat_region   21952. .22077

repeat_region   22083. .22259
                /note="L2 repeat: matches 2578. .2707 of consensus"
repeat_region   22280. .22393
                /note="FRAM repeat: matches -2. .175 of consensus"
repeat_region   22394. .22748
                /note="L2 repeat: matches 2541. .2645 of consensus"
repeat_region   22749. .22846
                /note="MLT1B repeat: matches 1. .390 of consensus"
repeat_region   22864. .23010
                /note="L2 repeat: matches 2645. .2745 of consensus"
repeat_region   23011. .23310
                /note="L2 repeat: matches 2329. .2429 of consensus"
repeat_region   23311. .23391
                /note="Alu3b repeat: matches 1. .299 of consensus"
repeat_region   23433. .23730
                /note="L2 repeat: matches 2429. .2521 of consensus"
repeat_region   23731. .25397
                /note="MSTD repeat: matches 66. .394 of consensus"
repeat_region   25477. .25556
                /note="L1PB1 repeat: matches 4442. .6123 of consensus"
repeat_region   25557. .25651
                /note="MSTD repeat: matches 1. .85 of consensus"
repeat_region   25745. .25785
                /note="L2 repeat: matches 2585. .2692 of consensus"
repeat_region   26898. .27373
                /note="L2 repeat: matches 2702. .2742 of consensus"
repeat_region   27378. .27483
                /note="MER31A repeat: matches 1. .661 of consensus"
repeat_region   27484. .27868
                /note="MER93 repeat: matches 292. .397 of consensus"
misc_feature    27537. .27688
                /note="match: GSS: Em:AQ127976"
repeat_region   27912. .28051
                /note="MER93 repeat: matches 1. .188 of consensus"
repeat_region   29900. .30121
                /note="L1PA13 repeat: matches 6011. .6152 of consensus"
repeat_region   30122. .30487
                /note="HERVL repeat: matches 3020. .3235 of consensus"
repeat_region   30488. .30578
                /note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region   30579. .30796
                /note="HERVL repeat: matches 2933. .3020 of consensus"
repeat_region   30797. .31887
                /note="MER58A repeat: matches 1. .221 of consensus"
misc_feature    complement(31888. .32090)
                /note="HERVL repeat: matches 1792. .2933 of consensus"
misc_feature    32922. .33453
                /note="match: GSS: Em:AQ759038"
                match: GSS: Em:B58153; match: GSS: Em:B66544;
repeat_region   33175. .33412
                /note="match: GSS: Em:AQ751315"
repeat_region   33454. .33885
                /note="L2 repeat: matches 2473. .2708 of consensus"
repeat_region   33896. .33946
                /note="MLT1C repeat: matches 6. .466 of consensus"
repeat_region   33947. .34302
                /note="MIR repeat: matches 162. .215 of consensus"
repeat_region   34303. .34334
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   34335. .34703
                /note="MIR repeat: matches 129. .162 of consensus"
repeat_region   34704. .34830
                /note="THE1C repeat: matches 1. .369 of consensus"
repeat_region   34831. .35552
                /note="MIR repeat: matches 20. .129 of consensus"
repeat_region   36612. .36657
                /note="L1MC2 repeat: matches 5573. .6325 of consensus"
repeat_region   36936. .37004
                /note="L2 repeat: matches 2675. .2723 of consensus"
repeat_region   37383. .37648
                /note="L2 repeat: matches 2638. .2707 of consensus"

Query Match    35.5%;  Score 488.2;  DB 9;  Length 124531;
```

Best Local Similarity 47.4%; Conservative 213; Mismatches 290; Indels 36; Gaps 2;
Matches 485; Pred. No. 1.9e-106;

```
Qy    389 argargaraathwsnaarcarcaarwsnathcaargargtnacntgggtnytnytnaargcnt 448
          | ::| |   ::  |  ::|::|| ||||:::|| :  :|||:| ||| || |  :|||||:|
Db  96092 AAGAAGGANTTCCTAAGCACGCAGCAAAGCATTTGAGATGTGACTTTGGTGTCTTGTTAAAGGCAT 96033

Qy    449 tywsnttyathmgngcngarcayaatwsnswsngaraayytncayccngayaayytna 508
          |::|| || | ||  :||| |  |||  :|:|: ||||  |:|  ::|| | |:  | |
Db  96032 TCAGTTTTATAAGGGAAGCAGAGAGCAATAAAAAATTTGCAGCCTGACTANGCAA 95973

Qy    509 thaararaaraaaycnttywsngarggnaarttyaaarytngcngcngarathtgyatht 568
          | :: |  |||||  || |::|| || ||  |||  ||: :||| ||| |||:||:||
Db  95972 TAGAAAAAGAAAAACCCATTTTCTGGGAGAAAACTCAAGCCTATTGCAGAAAATTTGCCTAA 95913

Qy    569 gyaaygargarytnaaygtnaayccncargayaaygyngaraayathwsntggacntgyc 628
          | ::| | | || |||||||::| :| ::|| || |::|| || :|  ||: :|||| :||
Db  95912 GTAGCAAGGAGCCTPAATGTTAATCCCCAAGACCATCGGGAAAGTCTPCACAGGCCATGTC 95853

Qy    629 armgnwsnwsncacaarwsnathaarwsnytngcntggmgnccnmgnmg---naartggt 685
          | :| | :  :|   || :||::|  | | || || |:||||  :   || ||||:|||
Db  95852 AGGGAGAACTTCATGGCTGCCCCTPCACATCACAAGCCTGCAGCGTAAAAAGTGGT 95793

Qy    686 tytgygygnacngngnccngnwsnytntgtygtgytncarccnmngayytngtnccntgyg 745
          | | :| | ||  |:|:|  |  ||||: ||: ||  |||| ::|| ||| |||||||:|
Db  95792 TTCAnGGGCCAGGCGTTCCTGTGTCCGTGTCCAGCACAGGGACTTGGTGTCCCTGTG 95733

Qy    746 tnccng-------------------------tnaaywsngcngtngcnwsng 772
          | ||:|                         |||| | |||||||||| | |
Db  95732 TCCCAGCTGCTCCAGCTATGGCTGAAAGGGGCCAAGGTAGAGGTTGGGCTTGGCTTCAG 95673

Qy    773 arggngcnwsnccnaaarccntggacngryytncccnwsnggngtngagcccngtngggngcnaaara 832
          | || |||:  :|  ||| |||||  |: |||:: :  | || ||: | ||| |:||| | ||||| |
Db  95672 AGGGTGCNAAGACCAAAGCCTTGGCAGCTTCCACGTGATPGTTGGGTGCACAGA 95613

Qy    833 arwsnmgnathgargtntgggaarccncncnaathmgnttycaararaaratthtaygngnaaycct 892
          | :|| |||:| ||||| ||||  || :| | ||| ||||:||||| | | ||||:||:||| ||||
Db  95612 AGTCAAGAACTGAGGTTGGGAACCTCTGCCTAGAATTCAGAAGATGTAATGGAAACGCCT 95553

Qy    893 ggatgccnmgncaaraaarttygcngtngggntngggnwsnwsntggmgnnacnwsncnmgng 952
          |||||||  || |||| ||:|||| |:||||| |||| | | :|||| |::|| :|  :|:|
Db  95552 GGATGCCCAGGCAAAAGTTTGCTCAGGGGTGTGGGGACCCTCATGGAGGAAAACTCTGCCTAGGG 95493

Qy    953 tngtncaaarggnaaaygtnggggtngggyaarccncncncaymgngtnccnwsnggngcnccnw 1012
          |:|| ||| || |||| ||||||| |||| ||| :| :| | ||:|||||| :|| ||| || |
Db  95492 CAGTGCGAAGGAAGAAAATGTGGGGTGCAGAGCCCPCCACCACAGAGTCCCTPCACTGGGGCACCAC 95433

Qy   1013 snwsnmgngcngtnmgnmgnwsncncncnwsnmgmgntncaraaynggnmgnwsnacng 1072
           :::| |:||||| || |:|| ||:| | :::| :|:| |||| ||:|| :|| | ||:
Db  95432 CTAGTGAGCTGTGAGCACPGGCCACTGYCCTCCAGACCCCCAGAATGATNGATCCACTG 95373

Qy   1073 aywsnsnytncarcaygtnccngaaraarwsnacngayacncaartgycarccngtnaargcng 1132
          | :|| ::|||| |||||:||||||||| :|||| | || |||| ||:|| ||:||||||||| |
Db  95372 ACAGCTTGCACCATTCACCTGGAAAAGCCACAGAGCACPCAAACGCCCATGAAAGCAG 95313

Qy   1133 cnggnatggarwsngtnccntayaaraacngtngtngcngarytnacnacacngtnggna 1192
          || || ||| |:||| |||||| |||| | ||:|:||| ||:|| || || |||:|||| 
Db  95312 CCAGGAGGAGGCTGTACCCTGCAAAGCCACACAGGGTGGGAACTGCCCAAGACTATGGGAA 95253

Qy   1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgnaycayttyg 1252
          || || || ||||| || ||| | || |||:|||||  || ||||| ||:|| || | |
Db  95252 CCTACCTCTPGCCATCAGCATGACCTGGATGTGTGTCAAAGGAGGAGATCATTPTTG 95193

Qy   1253 gngcnytnmgnttygaytgyccnacngngtnacntayatgggacntnccnytnt 1312
          ::|| |||:|||||| ||:| ||| |:||||| ||| :|||| ||| ::|||:|
Db  95192 AAGCTTTAAAATTTGACTGCCTCACTGGAATTTCATGGCCCTTGTACCACTPT 95133

Qy   1313 gyttyggncnarttyttyccnttyggnacngcngtntttyacncaartgyytntayytncayt 1372
          | || |||||| || || |||||||| | |||: ::||:|| ||| | ||: |||| ||| |
Db  95132 GTTTTGGCCAANTTCTCCCATTTGGGAGGGCTATATTTACCCCAAATACTTGTACCCCCATT 95073

Qy   1373 gyat 1376
          | ||
Db  95072 GTAT 95069
```

Db 95072 GTAT 95069

RESULT 14
LOCUS   HSU212C1   40714 bp   DNA   linear   PRI 23-NOV-1999
DEFINITION Human DNA sequence from cosmid U212C1, between markers DXS87 and
            DXS356 on chromosome X.
ACCESSION   Z69722
VERSION   Z69722.1 GI:1204119
KEYWORDS   X.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 40714)
  AUTHORS   Odell,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-FEB-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquiries: humpub@sanger.ac.uk
COMMENT   IMPORTANT: This sequence is the entire insert of clone U212C1. The
            true left end of clone U212C1 is at 1 in this sequence. The true
            right end of clone U212C1 is at 40714.
            U212C1 is from the human chromosome X-specific cosmid library.
FEATURES   Location/Qualifiers
     source   1..40714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="X"
                     /clone_lib="LL0XNC01"
                     /clone_lib=LL0XNC01-212C1"
     repeat_region   10..233
                     /note="L1 element fragment"
     repeat_region   388..758
                     /note="L1 element fragment"
     repeat_region   932..1108
                     /note="L1 element fragment"
     repeat_region   2571..2617
                     /note="MER43 element fragment"
     repeat_region   2613..5203
                     /note="L1 element fragment"
     repeat_region   5204..5229
                     /note="13 copies of 2 mer 100 % conserved"
     repeat_region   5230..5427
                     /note="MER43 element fragment"
     repeat_region   8954..9088
                     /note="L1 element fragment"
     repeat_region   12171..13985
                     /note="L1 element fragment"
     repeat_region   14179..14612
                     /note="L1 element fragment"
     repeat_region   14662..14780
                     /note="L1 element fragment"
     repeat_region   14824..15044
                     /note="L1 element fragment"
     repeat_region   15234..15369
                     /note="L1 element fragment"
     repeat_region   15421..15849
                     /note="L1 element fragment"
     repeat_region   16026..16155
                     /note="MSTA element fragment"
     repeat_region   16073..16149
                     /note="THE1B element fragment"
     repeat_region   16204..16495
                     /note="Alu repeat: matches 1..308 of consensus"
     repeat_region   16532..16609
                     /note="THE1B element fragment"
     repeat_region   16624..16722
                     /note="MSTC element fragment"
     repeat_region   16626..16737
                     /note="MSTA element fragment"
     repeat_region   16651..16712
                     /note="MSTA element fragment"

```
repeat_region   17060..17653
                /note="THE1B element fragment"
repeat_region   19720..19797
                /note="L1 element fragment"
repeat_region   19877..19915
                /note="MER21B element fragment"
repeat_region   20007..20256
                /note="MER21B element fragment"
repeat_region   20518..20734
                /note="MER21B element fragment"
repeat_region   22661..22730
                /note="MER30 element fragment"
repeat_region   22740..22814
                /note="MLT1B element fragment"
repeat_region   22917..22974
                /note="MLT1A element fragment"
repeat_region   23342..23501
                /note="MLT1A element fragment"
repeat_region   23550..23697
                /note="THE1B element fragment"
repeat_region   23725..24279
                /note="THE1B element fragment"
repeat_region   24495..24593
                /note="THR element fragment"
repeat_region   24643..25270
                /note="THR element fragment"
repeat_region   25277..25463
                /note="THR element fragment"
repeat_region   25528..25627
                /note="THE1B element fragment"
repeat_region   26081..26196
                /note="THE1B element fragment"
repeat_region   26265..26661
                /note="MLT1A element fragment"
repeat_region   26797..26898
                /note="LTR2 element fragment"
repeat_region   27949..28241
                /note="MSTC element fragment"
                /partial
repeat_region   28319..28608
                /note="Alu repeat: matches 308..1 of consensus"
                /partial
repeat_region   30032..30159
                /note="Alu repeat: matches 308..1 of consensus"
                /partial
repeat_region   31794..31980
                /note="MER5 element fragment"
                /partial
repeat_region   31981..32073
                /note="Alu repeat: matches 308..109 of consensus"
                /partial
repeat_region   33232..33735
                /note="Alu repeat: matches 93..1 of consensus"
repeat_region   35168..35654
                /note="MER9 element fragment"
repeat_region   40640..40693
                /note="MER4B element fragment"
                /note="27 copies of 2 mer 94 % conserved"

BASE COUNT      11544 a     7879 c     8141 g     13150 t
ORIGIN
```

```
Qy    334  ytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaayaaargar  393
           |:|:|||: |:|||::|||: ||:|||||::||| ||||||||: |:|||||::||
Db  24003  CTGCCCTAGAGATCTGGAATTGACCTTGAGAGAGATGAGTTTAGGGTATCAGGTGGAA  24062

Qy    394  garathwsnaaarcaarwsnahcaracaargtnacntcgggytnnytnaaargcnttywsn  453
           ||:||:::: ||:||:|::|| ||::||:|:||:|||:||||:|||||||||:|||||::
Db  24063  GAAATTTCTAAGCAGCAAAGCAAAGCAATTCAAGAGAGGTGACTTGTGTCCTGTTAAAGGCACTCAGT  24122
```

```
Qy    454  ttyathmgnargcngarcayaarwsnwsngaraayytncaycccngayaaygtnathaar  513
           ::|:||||||||: |:||||||||::|||:||||||||||||||::|: |||||||||:|
Db  24123  TTTGCAAGGGAGGCAGAAACAATAAAAAGTTCAGAAAAATTTTCACCCTGACAAGTGATAGAA  24182

Qy    514  aararaaaycc-nttywsngarggnaaarttyaaarytngcngncgararathtgyathtgyaa  572
           |||||||:|| |::||||||||||: ||:||:||||||||||||||:|||||: |||:
Db  24183  AAGAAAAACCCATTTTTCTGATGATAAAATTGAAGCCAGCGTGCAGAAAATTTGCATAAGTAA  24242

Qy    573  ygargarytnaayggtnaaycccacargayayaygngaraayathwsntggacntgycarng  632
           :|| |||| |:||:| |||: |||:||||| |||| ||:|||| ||:| | |:||| |
Db  24243  CGAGGAGCCAAATGTTAATCACTAAGACCATAGGGGAAAAATATCTCCAATGGCTTGAGAG  24302

Qy    633  nwsnvsnscarcarwsnaathaarwsnyt-ngcntgymgnccnmgnmgnnaattggtytytgyg  691
           :||:|||:||||| |||||||||||:| |||:|||||||: |||||| |||||||||||:|:|
Db  24303  GTCTTCACGGCAGCCCCTCACTCACTGCCCCCAGACCCTAGGAGAAAAATGGTTTTATG  24362

Qy    692  gnacnggnccnggnwsnytntcgytgyngtncarccnmngnayytngtcnccntgyygtnccng  751
           ::||||||| |||| :|:||:||||:|| |:||| ||:|||::| |:||||::|:|:||
Db  24363  GGTCGGGCGCCCAGGGTCCCAATGGCTTCCCATGCGTATGTGCAGTCTRAGGGACTGGTGCCCCGTGTCCCAG  24422

Qy    752  tnaa--------------------ywsncngtngcnwsnargggng  778
                                    ::||||||||||||:|||:|:|
Db  24423  CCACTCCAGCCATAACTGAAAAGGGGGCCCAAGGTGTATACCTTGGGCTTGCTTCAAAGGATG  24482

Qy    779  cnwsnccnaarccntggcarytnccnwsnggngtngarccncgntgngngcnaararwsnm  838
           |: |||||| |||||||||:||:||||:||||||| ||:|:||: ||:|: ||||||::
Db  24483  GAAGCCCCAGGCCTTGGCGGCTTCCCATGTGGTGTTGAGCCTGCACTGAAGTCAA  24542

Qy    839  gnathgargtntgggarccnccnaathmgnttycaraaaarathtaygngnaaycccntggatgc  898
           |:||||||||||||||:||:|||:||:|||||||:|:|||| ||:||||:||||:||||||||
Db  24543  GAATTGAGGTTTCGGGAACCTCCACTGARTTTCAGAAGAGTGTATGGAAAATGCCTGGAATGC  24602

Qy    899  cnmgncaraarttygcngtnggngtnggnwsntggmgnacnwsngcnngngtagtnc  958
           ::|||: | |::||| ||||| |:|| :|||||| ||| ||:|||:||:|||||||:
Db  24603  CCAGGCAAAAAGTTTGCTGCAGGGGCGGAGGAACCCTCGTGTTRAGGTGGTGGT  24662

Qy    959  araarggnaaygtngngtnggarccncncaymgntgcncnwsnggngcncccnwsnwsnm  1018
           :||||||| | :|||||||||:||:|||| ::|||||:|||||||||:::|||||:::
Db  24663  GGAAGGGGAAAAAGTGGGGTCATTAGCCGCCACACAGAGTCCCTACCAGGGACCACCTAGTG  24722

Qy   1019  gngcngngtnmgngmgwsnccncccnwsnwsnsngngnytncaraarggnmgnwsnaacngayawsny  1078
           |||:||:||:|||||||:||::|:|| |:||||| |:|:|||||||:::||:|||||||::|:
Db  24723  GAGCTGTGAGAAGAGAGGGCCCCACCATGTCTCCAGACCCCAGAGATGGTAGATCACCGAGACAGCC  24782

Qy   1079  tncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngngna  1138
           ||:|| |||||:||||||| |:||| ||| |:||:|||: |:||| ||||: |||| |||
Db  24783  TGCACCGTGTGCCTGGAAAAAGCCGCAGACACTCAATGCACGACCTGTGAAAACGAAGGGA  24842

Qy   1139  tggarwsngtnccntayaaracngtngcngaytnacnaaracngtnggnnathtayy  1198
           |||||||:|||:|||:||:|||:||||||||||::| |||::|||||| ||:||||
Db  24843  GGGAGGCTGTACCCTGCAAAGCCACAGGGCGAGCTGCCCAAGATCCATGGGAACCAATC  24902

Qy   1199  tnytncaytgycayggayytngayytnmgncayggngtnaarmgnagaycayttyggnngcny  1258
           :|:|| |||||: ||:||||: ||:||| ||:|||:|||||| |||| ||:|||||||:|||
Db  24903  TCTTGCATCAGACCTGGATATGAGACAATGGAGTCAAAGGAAATCATTTTGGAGCTT  24962

Qy   1259  tnmgnttygaytgycnacnggnttymgnacntayatgggngnccngt-nccnytntgytty  1317
           ::|:||||||||||::|||:|||::|| ||:|||||||||| ||||| ||:|:|||:||:|
Db  24963  TAAGATTTGACTGTCCCATGGCTTCAGACTTTCCATGGGCGCCTTCGTTTTT  25022

Qy   1318  ggncarttytyccnttyggnacngcngtnttyacncaartggytntayytncaytgyat  1376
           ||:|| ||:||:|::|||||:|||||||:||:|| ||||||| |||:||||:|||||||
Db  25023  GGCCAATATCTCCCATTTGAATGGCTGTATTATCCAATGCGTACCCCATTGTAT  25081
```

```
SOURCE       Homo sapiens pre-pro-B cell cell_line:FLEB 14 - 14 DNA,
             clone_lib:Keio BAC library clone:KB1222D11.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 159712)
  AUTHORS    Shimizu,N. and Asakawa,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-SEP-2001) Nobuyoshi Shimizu, Keio University, School
             of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
             160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
             Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT      On Jan 14, 2002 this sequence version replaced gi:15824050.
FEATURES             Location/Qualifiers
     source          1..159712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1222D11"
                     /cell_line="FLEB 14 - 14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     repeat_region   1..4242
                     /rpt_family="L1PA2"
     repeat_region   4668..4733
                     /rpt_family="MIR"
     repeat_region   5899..5921
                     /rpt_family="(TGGA)n"
     repeat_region   complement(6205..6406)
                     /rpt_family="MIR"
     repeat_region   6410..6589
                     /rpt_family="FRAM"
     repeat_region   6619..6648
                     /rpt_family="AT_rich"
     repeat_region   complement(6649..6776)
                     /rpt_family="AluJo/FRAM"
     repeat_region   6860..7158
                     /rpt_family="AluJo"
     repeat_region   9083..9127
                     /rpt_family="AT_rich"
     repeat_region   9188..9226
                     /rpt_family="(TA)n"
     repeat_region   complement(9227..9522)
                     /rpt_family="AluSq"
     repeat_region   9836..9933
                     /rpt_family="L2"
     repeat_region   10728..10762
                     /rpt_family="L1MA1"
     repeat_region   10763..11062
                     /rpt_family="AluSg"
     repeat_region   11063..11300
                     /rpt_family="L1MA1"
     repeat_region   11301..11324
                     /rpt_family="AT_rich"
     repeat_region   11520..11545
                     /rpt_family="AT_rich"

                     /evidence=not_experimental
                     12507..12560
                     /rpt_family="(TAAAAA)n"
                     /evidence=not_experimental
                     12793..12825
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
                     15792..15852
                     /rpt_family="AluJ/FRAM"
                     /evidence=not_experimental
                     15993..16013
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
                     complement(16014..16312)
                     /rpt_family="AluSq"
                     /evidence=not_experimental
                     17126..17436
                     /rpt_family="AluSg1"
                     /evidence=not_experimental
                     18198..18489
                     /rpt_family="AluSc"
                     /evidence=not_experimental
                     18490..18715
                     /rpt_family="L2"
                     /evidence=not_experimental
                     complement(18801..18936)
                     /rpt_family="FLAM_C"
                     /evidence=not_experimental
                     complement(19119..19294)
                     /rpt_family="L2"
                     /evidence=not_experimental
                     19346..19541
                     /rpt_family="L2"
                     /evidence=not_experimental
                     19687..19844
                     /rpt_family="L1MEc"
                     /evidence=not_experimental
                     complement(19954..20081)
                     /rpt_family="L2"
                     /evidence=not_experimental
                     20082..20106
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
                     20423..20566
                     /rpt_family="FRAM"
                     /evidence=not_experimental
                     20567..20636
                     /rpt_family="GA-rich"
                     /evidence=not_experimental
                     complement(20973..21282)
                     /rpt_family="AluY"
                     /evidence=not_experimental
                     22379..22724
                     /rpt_family="L2"
                     /evidence=not_experimental
                     22933..23019
                     /rpt_family="L2"
                     /evidence=not_experimental
                     23044..23270
                     /rpt_family="L2"
                     /evidence=not_experimental
                     complement(23275..23586)
                     /rpt_family="AluSg1"
                     /evidence=not_experimental
                     complement(23946..24035)
                     /rpt_family="MER94"
                     /evidence=not_experimental
```

```
repeat_region    24491..24777
                 /rpt_family="Aluy"
                 /evidence=not_experimental
repeat_region    24782..24794
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    24795..24909
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    24910..24939
                 /evidence=not_experimental
                 /rpt_family="L1PA4"
repeat_region    complement(28598..29054)
                 /evidence=not_experimental
                 /rpt_family="AluSp"
repeat_region    complement(27879..28173)
                 /evidence=not_experimental
                 /rpt_family="AluSp"
repeat_region    complement(24981..25276)
                 /evidence=not_experimental
                 /rpt_family="AluJb"
repeat_region    30710..30738
                 /evidence=not_experimental
                 /rpt_family="MER63B"
repeat_region    complement(30739..30993)
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    30994..31032
                 /evidence=not_experimental
                 /rpt_family="AluJb"
repeat_region    complement(31408..31724)
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    complement(31949..32254)
                 /evidence=not_experimental
                 /rpt_family="AluJo"
repeat_region    32336..32362
                 /evidence=not_experimental
                 /rpt_family="AluSg"
repeat_region    complement(32463..32565)
                 /evidence=not_experimental
                 /rpt_family="AT_rich"
repeat_region    32685..32898
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    32995..33020
                 /evidence=not_experimental
                 /rpt_family="MIR"
repeat_region    33995..33020
                 /evidence=not_experimental
                 /rpt_family="(TAAA)n"
```

```
Query Match          35.4%;    Score 487.6;  DB 9;    Length 159712;
Best Local Similarity  48.0%;   Pred. No. 2.9e-106;
Matches  484;  Conservative  210;  Mismatches 293;  Indels 22;  Gaps 3;
```

```
Qy     390  rgargarathwsnaarcarwsnathcargargtnacntgggtnytnaytnaargcntt  449
            |||:||| :| ||: |    : |||||  |||||||  |  ||||||||||||||||
Db  138232  AGAGAAGTTCCTAAGCACAAAGCCATTCAAGAGAGATAACTTGGGTGTCATTAAAGGCACT 138173

Qy     450  ywsnttyathmgngargcngarcayaarwsnwsngaraayytncaycngayaaygtnat  509
            ::: |||:|||:|||||||||||| |||||||||||||||||||||||||||| ||||
Db  138172  CAGTTTTATAAGGGAAGCAGAGCATAAAACTTTGGAAAANTTCGCAGCCTGACTATGCAAT 138113

Qy     510  haaraaraaraayccnttywsngargngnaarttyaarytyncgncngarathtgyathg  569
            :|| |||:||||||||||||| |||| |||:|||||||| || ||||:|||:||:|
Db  138112  AGAAAAGAAAAAANCCCATTTTCTGGGGAGAAAATTCAAACCGGCTGCAGAAAATTGCACAAG 138053

Qy     570  yaaygargaryrytnaaygccncargayaaygngaraayathwsntggacntgyca  629
            :| |:|||| ||||| |||: |||||:||||:|||||||||:|||||||||:|||
Db  138052  TAGCAAGGAGGCCAAAATGTTAATCCCAGGACCATGGGGGAAAAAGTTCTCCAGGCCATGTCA 137993

Qy     630  rmgnwswsncacarwsnathaarwsnytngcntgggngnc---cnmgnmgnaartggtt  686
            :||:| |||||  |  ||||||||||:||| |||:||||||||||||||||||||||||
Db  137992  GAGACCTTTACAGCAGCCCCTCCATCATTAAGCCTGGAGGCCCTACGAGGGAAAAGTGGTT 137933
```

```
Qy     687  ytgyggnacngngngggnwsnaytntgyttgygtncaarccnmngngayyytngtncntgygt  746
            || |:||| |||  |||| ::||: | : : ||:|||||||| :||  |:||  |||:||
Db  137932  TCATGGGCGCCTGGCCCCAGGGTCCCTGGGGCTGTGTGCAGCCTAGGGACTTGGTGACCTGTGT 137873

Qy     747  nccngtnaaywsngcngtngcnwsngargg------------------ngcnwsnccnaa  788
            |||||  :: |||:|||||| ||||:||                      ||| |  |||
Db  137872  CCCAGCCACTCCAGCCGTGCCTGCTGAAAGGGGCGACTGTACAGCTTGGGCTTGTGGCTTCAGA 137813

Qy     789  rccntggcaarytnccnwsnggngtnggnarccngtnggngcnaaraarwsnmgnaathgargt  848
            : |||| ||| ||||| | ||||||| |||||||||||:|| |||||| |:|||:|||||
Db  137812  GGCTTGGCAGCTTCCACCTGGTTGTTGACGCTGTTGGTGCACAGAAGTTCAAGAATTGAGGT 137753

Qy     909  rttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaaraaggnaa  968
            :||:|||| ||||| ||  : ||:| | |||||| :|||| |: ::| |||||||||||
Db  137752  TTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAATCCCTGGATGCCAAGCCAAAA 137693

Qy     909  rttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaaraaggnaa  968
            :||:|||| ||||| ||  : ||:| | |||||| :|||| |: ::| |||||||||||
Db  137692  GTTTCTGCGCAGGGGTGGGGGGCCCCTCATTGCTGAGAAACCTCTGCTAGGGCAGTGAGGAAGGAAA 137633

Qy     969  ygtnggntgggarccncnccncaymgntgtnccnwsnggngcnccnwsnwsnmgngcngtnmg  1028
            ::|| :||||| || ::||  |    | |  |:||||| ||||  : || | :|||||| ::
Db  137632  TGTGGGGGTTCGATTCCCCACACAGAGTCCCTACTGTGGGCACTGCCTAGTGGGAGCTGTGAG 137573

Qy    1029  nmgnwsnccnccnccnwsnwsnngnytncararggngmgnwsnaacngayawsnnytncaarcaygt  1088
            | | | :||| | |  :| | | : || ||||:| :| |||||| | :|:|| | |||||||
Db  137572  AAGAGGGCCACCATCCTTCAGACCCCAGAATGGTAGATCCACTGACAGCTTGCACCACCATGT 137513

Qy    1089  nccngararwsnacngayacncaartgycarccngtnaargcngcnggngatggtgargar wsngt  1148
            ||| |::| | |||:|||| :|::  ::|| |||| |||:| |||:||| | |: || |:| ::|
Db  137512  GCCTGAAAAACTGCAGACACATTCAATGCCCCAGCCCATGAAAAGCAGCAGGAGGAAGGCTAT 137453

Qy    1149  nccntayaaracngtngtngcngarytnacnaaracngtnggnaathtayytnytncaaytg  1208
            | ::|| ||  ||| ||:|| |||| ||| |||| |||:|||| | ||||| || |||:||
Db  137452  ACCCTCCAAAGCCACAGGGCGTGCCCAAGACCACCTACCTTTTGCCATCA 137393

Qy    1209  ycayyayytngaygtnmgncayggngtnaarmgnaaycaytygngcnytnmgnttyga  1268
            :|| |||||||||| | |||||| ||:|| ::|||| || |::||| |:| :|||||
Db  137392  GCGTGACCTGGATGTGAGACCTGTAATCAAAGGAGATCATTTTGGAGCTTTAAAATTTGG 137333

Qy    1269  ytgyccnacngngnttymgnacntayatgggnaccngt-nccnytnttgyttyggncarttyt  1327
            :||||| |||| || ||:| |||| |||||| ||||| |||||:|||| || :|||:| |||
Db  137332  CTGCCCACCACTTGTTGTTTGGACTTCCATGGGCCCTGTAAACCTCTTTGTTTTGGCTAATTTT 137273

Qy    1328  tyccnttyggnacngcngtnttyacncaartgyytntayytncaytgyat  1376
            |:|| ||||: ||||||| | ||| |||||| ||:| ::|| |||||||
Db  137272  TCCCATTTGGAACAGTAGCGTTTCCCCAAATACCTGTACTCCCATTGTAT 137224
```

```
Search completed: July  8, 2002, 18:08:58
Job time: 18594 sec
```

```
GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:               July 8, 2002, 12:53:47 ; Search time 481.64 Seconds
                                               (without alignments)
                                               4908.627 Million cell updates/sec

Title:                US-09-997-610-3
Perfect score:        1377
Sequence:             1 athgtngtnathccngtnyt.........gyytntaytncaytgyatg 1377
Scoring table:        IDENTITY_NUC
                      Gapop 10.0 , Gapext 1.0
Searched:             1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters:       3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :        N_Geneseq_032802:*
    1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
    2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
    3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
    4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
    5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
    6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
    7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
    8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
    9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
   10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
   11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
   12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
   13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
   14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
   15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
   16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
   17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
   18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
   19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
   20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
   21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
   22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
   23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
   24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
                                    %
Result                           Query
No.    Score   Match  Length  DB  ID         Description
------ -----   -----  ------  --- --         -----------
   1   471.4   34.2    2590   22  AAH14327   Human cDNA sequenc
   2   452     32.8    3093   22  AAS03687   Rhesus gene locus:
   3   452     32.8    3093   22  AAS03689   Rhesus gene locus:
   4   452     32.8    9236   22  AAS03690   Rhesus gene locus:
   5   452     32.8    9241   22  AAS03688   Rhesus gene locus:
   6   429.8   31.2    3203   22  AAD05134   Human secreted pro
   7   427.6   31.1    6063   22  AAL03635   Human reproductive
   8   427.2   31.0   22509   22  AAS44505   Human LEKTI DNA c1
   9   427.2   31.0   31529   22  AAS44506   Human LEKTI DNA c1
```

ALIGNMENTS

```
RESULT     1
ID    AAH14327
AC    AAH14327;
      AAH14327 standard; cDNA; 2590 BP.
XX
DT    26-JUN-2001  (first entry)
XX
DE    Human cDNA sequence SEQ ID NO:11697.
XX
KW    Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS    Homo sapiens.
XX
PN    EP1074617-A2.
XX
PD    07-FEB-2001.
XX
PF    28-JUL-2000; 2000EP-0116126.
XX
PR    29-JUL-1999; 99JP-0248036.
PR    27-AUG-1999; 99JP-0300253.
PR    11-JAN-2000; 2000JP-0118776.
PR    02-MAY-2000; 2000JP-0183767.
PR    09-JUN-2000; 2000JP-0241897.
XX
PA    (HELI-) HELIX RES INST.
PI    Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI    Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR    WPI; 2001-318749/34.
```

```
   10   426.4   31.0    6063   22  AAL03634   Human reproductive
   11   421.4   30.6    1736   22  AAH13678   Human cDNA sequenc
        421     30.6  465237   24  ABA90193   Human oestrogen re
   12   409     29.7    1485   23  AAS91984   DNA encoding novel
   13   408.8   29.7    1278   22  AAD05178   Human secreted pro
   14   400.4   29.1    1580   23  AAS87262   DNA encoding novel
   15   395.2   28.7    7726   22  ABA21079   Human nervous syst
   16   393.6   28.6   32127   22  AAI99255   Human excretory re
   17   393.6   28.6   32127   22  ABA163605  Human kidney relat
   18   393.6   28.6   72215   22  AAK86832   Human excretory re
   19   388.2   28.2    1405   22  AAK70265   Human immune/haema
   20   388.2   28.2    1405   22  AAK70266   Human immune/haema
   21   386.6   28.1    8387   22  ABA14892   Human nervous syst
   22   386.4   28.1    8387   22  AAK84548   Human immune/haema
   23   386.4   28.1    1621   23  AAS67924   DNA encoding novel
   24   384.4   27.9    1621   23  AAS64678   Human cDNA sequenc
   25   378.8   27.5    3049   22  AAS97997   DNA encoding human
   26   373.8   27.1   81369   21  AAA16637   DNA encoding novel
   27   371.6   27.0    1306   23  AAS40321   Human T gene DNA.
   28   361.6   26.3   22756   22  AAS64321   Human reproductive
   29   361.6   26.3   22756   22  AAL03921   Human reproductive
   30   355.4   25.8    1946   22  ABA44227   Human breast cell
   31   355.4   25.8    1946   22  ABA54677   Human foetal liver
   32   355.4   25.8    1946   22  ABA24459   Probe #2925 for ge
   33   355.4   25.8    1946   22  AAK02964   Human brain expres
   34   355.4   25.8    1946   22  AAK28410   Human bone marrow
   35   355.4   25.8    1946   22  AAI12975   Probe #2908 for ge
   36   355.4   25.8    1946   22  AAI34334   Probe #3020 used t
   37   355.4   25.8    1946   22  AAI02893   Probe #2884 used t
   38   347     25.2   26410   22  AAK70623   Human immune/haema
   39   346.8   25.2   32986   22  AAK69758   Human immune/haema
   40   346.8   25.2   32986   22  AAK84629   Human immune/haema
   41   346.4   25.2     923   22  AAS66623   DNA encoding novel
   42   341.4   24.8   50000   24  AAD26400   Human glutamate re
   43   341.4   24.8   50000   24  AAD26437   Human GRM3 gene fr
   44   337.4   24.5   57728   22  AAC87588   Human 9p11 chromos
   45   336.6   24.4    2197   22  AAH18389   Human cDNA sequenc
```

```
XX
PT  Primer sets for synthesizing polynucleotides, particularly the 5602
PT  full-length cDNAs defined in the specification, and for the detection
PT  and/or diagnosis of the abnormality of the proteins encoded by the
PT  full-length cDNAs -
XX
PS  Claim 8; SEQ ID 11697; 2537pp + CD ROM; English.
XX
CC  The present invention describes primer sets for synthesising 5602
CC  full-length cDNAs defined in the specification. Where a primer set
CC  comprises: (a) an oligo-DT primer and an oligonucleotide complementary
CC  to the complementary strand of a polynucleotide which comprises one of
CC  the 5602 nucleotide sequences defined in the specification, where the
CC  oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC  of an oligonucleotide comprising a sequence complementary to the
CC  complementary strand of a polynucleotide which comprises a 5'-end
CC  sequence and an oligonucleotide comprising a sequence complementary to a
CC  polynucleotide which comprises a 3'-end sequence, where the
CC  oligonucleotide comprises at least 15 nucleotides and the combination of
CC  the 5'-end sequence/3'-end sequence is selected from those defined in
CC  the specification. The primer sets can be used in antisense therapy and
CC  in gene therapy. The primers are also useful for synthesising polynucleotides,
CC  particularly full-length cDNAs. The primers are useful for the
CC  detection and/or diagnosis of the abnormality of the proteins encoded by
CC  the full-length cDNAs. The primers allow obtaining of the full-length
CC  cDNAs, easily without any specialised methods. AAH03166 to AAH13628 and
CC  AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC  AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC  represent oligonucleotides, all of which are used in the exemplification
CC  of the present invention.
XX
SQ  Sequence 2590 BP; 722 A; 526 C; 670 G; 672 T; 0 other;

Query Match         34.2%;  Score 471.4;  DB 22;  Length 2590;
Best Local Similarity  47.0%;  Pred. No. 1.1e-108;
Matches 481;  Conservative 213;  Mismatches 293;  Indels 37;  Gaps 3;

Qy   390 rgargararthwsnaarcacarwsnathcarrgargtnacntggtgntytnaargcntt 449
         :|||.|||.||||.::||||.||.|||||||:||.||:|.||||.:.||:|.:||||.|
Db   608 agaagaaatttctaagcacgcaaagcaattcaagaggtgagttgggtactgttaagcatt 667

Qy   450 ywsnttyathmgnaargcngarcayaarwsnswsngaraayyyccngayaaygtnat 509
         :|.||:|.|||.||:|||.||.||||:.||||.||.|||||.||||.||.||:|.|
Db   668 cagttcaaaaggaaacagagcataaaaattcagaaaattgtagctgactatgtcaat 727

Qy   510 haararaaraayccnttywsngargngaarttyaarytngcngnaraththgyatthg 569
         :.|||.||||.||.||.||:||.|.||||||||||.||.|||.|||.||||.|||||
Db   728 agaagaaaaacaagtttctgggagaattcaagccagctgcagaaatctgcataag 787

Qy   570 yaaygaargarytnaaygtncncacaryayaaygngaraaayathwsntggacntgyca 629
         :.|:.|||.||.||:||.||.|:||||||||:.||.||||.|||.||.||||.|||||
Db   788 catcaagaagcctaatgttaatcccacatgggaaatgtctccaggccatgtca 847

Qy   630 rmgnwswsncarcarwsnathaarwsnytngcntggmgnccnmgnmg--naartggtt 686
         |.||.||:||||.|||.|||.||:.:||.|.|||||.||.|||.||.:.:|||.||||
Db   848 gagaccttcacagcagcccctcctgtcaggccaggatccaggagaaaatggtt 907

Qy   687 ytgyggnacngngncngnwsnytntgytgyngtncarccnmgngayytngtnccntggyt 746
         :||.||.|||.|.||.|.:|:.||||.||.||.||||.||.||.|||.||:|.|||||
Db   908 ttatggccaaggaggtcccctgctgctgtgtgcagcctatagactggtgccctgt 967

Qy   747 ncncngtnaayws----------------ngcngtngcnwsnga 773
         |.|:.||.|||.|              |.|.||.||:||||.|
Db   968 cccagctgccccagccatgtgaaaggccaatgtagagctcatgctgtggcttcaga 1027

Qy   774 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtngngcnaaraa 833
         |.||.||:||.||.||.|||||.||:|.||.||:||.||.||.||.|.|.||||||||
Db  1028 gggtggaagcccaagccttggcagcttccacatggtgttgagcctgtaggtacacagaa 1087

Qy   834 rwsmsmgnathgargtntggargcacnccnatmgnttycaraarathtayggnaayccntg 893
         |.:.:.:||.|||||||||||.||.||.||.|:.||||||||.||.|||||.|||.|||
Db  1088 gtcagaattgaggtttgggaacctccacttcagaagatgtatggaaatgcctg 1147

Qy   894 gatgccnmgncararttygcngtngntgggtggnwsnwsntggmgnacnwsngcnmgngt 953
         |||.|:.||.|||.|||||.|:.||.|:.||||||.||.||.||.||.||:||||.||.|
Db  1148 gatgtccaagtcaaaagtttgctcagtggaagggccctcatgggaagaacctctgtaggc 1207

Qy   954 ngtncaararggnnaaygtngntgggccggagctcccacgtaggtcctaatggggcattatc 1013
         |.||.||||.||:.||.|||.||||.||.||||||.|||.|.||:|.||.|||||.|||.|
Db  1208 aatgcagaaaggaattgtgggccggagctcccacgtaggtctctaaatgggcattatc 1267

Qy  1014 nwsnmgncngtnmgnmgnwsnccncncnwsnsmgnytncaraarggngmgnwsnacnga 1073
         |:.:.|.|.||:.||:||:.||.|.|.:||:.||.||||||.||.|:.||:||.||.||
Db  1268 tagtggagctcgtgagaagaggtcaccgttcctcagaccccagcagtgttggatccactga 1327

Qy  1074 ywsnytncaracaygtnccngaraarwsnacngayacncaatgycarccngtnaargcngc 1133
         :.||:|||||.||.||.||.||||||.||.||.|||.||.||:||.||.||:||||||.||
Db  1328 cagctcgcacatcgcctggaaaagccgcagacactcaatgcyatccgtgaaaagcagc 1387

Qy  1134 nggnatggarwsngtnccntayaaracngtngcngaryytnacnacracngtnggnnat 1193
         :||.||||||.||.||.||.|||||.||.||.||:.||.||.||.||.||.||::.||
Db  1388 caggaggagctgtaccctgcaaggccacagggggagaactgccaaagccatgggtcc 1447

Qy  1194 htayytnytncaytgycaygayytngayytnmgncaygyggngtgtnaarmgngaycayttygg 1253
         :.||:|.||.||||||||.||.|||.|||:.||.||||.||.:|||||||.:.|||||.:||||
Db  1448 ccacctttgcatcacgcctgaactgacacagatggagctctaggagagagatcatttgg 1507

Qy  1254 ngcnytnmgnttygaytgyccnacnggnttymgnaacntayatggnccng-tnccnytnt 1312
         :||.||:.||.||.||||||.||.||:.|||.||.|||.|||.||.||||  |.||||.|
Db  1508 agcttaaaatttgattgtcctgctggatttggacttgcatgggtcctgcagccccctt 1567

Qy  1313 gyttyggncarttytyccnttyggnacngcgtntttyacncaartgyyytntaayytncayt 1372
         |:||||||.||:||||.||:||||||.||.||.|.:.||.||.|.:.||.||.|||:.||:
Db  1568 gttttggccaatgtcccaatttggaatggctgtattaccaataccctgtaccctatc 1627

Qy  1373 gyat 1376
         :|||
Db  1628 gtat 1631
```

```
RESULT 2
AAS03687/c
ID  AAS03687 standard; DNA; 3093 BP.
XX
AC  AAS03687;
XX
DT  29-AUG-2001 (first entry)
XX
DE  Rhesus gene locus: RHD gene deletion in Rh negative haplotypes.
XX
KW  Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD;
KW  RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW  haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
OS  Homo sapiens.
XX
FH  Key             Location/Qualifiers
FT  primer_bind     32..54
FT                  /*tag= a
FT                  /note= "Binding site of primer rez7"
FT  primer_bind     complement (3034..3054)
FT                  /*tag= b
FT                  /note= "Binding site of primer rnb31"
XX
PN  WO200132702-A2.
XX
PD  10-MAY-2001.
XX
PF  31-OCT-2000; 2000WO-EP10745.
XX
PR  02-NOV-1999; 99EP-0121686.
PR  31-MAY-2000; 2000EP-0111696.
XX
PA  (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
```

PI
XX
DR   WPI; 2001-291052/30.
XX
PT   RHD positive haplotypes in D-negative
PT   and RHCE genes -
XX

Flegel WA, Wagner FF;

New nucleic acid molecular structure, useful for detection of common
RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
and RHCE genes -

Example 10; Fig 5; 135pp; English.

CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   RHD gene deletion in the Rh negative haplotypes. The Rhesus genes
CC   locus comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RHD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;

Query Match   32.8%;   Score 452; DB 22;   Length 3093;
Best Local Similarity 47.4%;   Pred. No. 1.1e-103;
Matches 485; Conservative 212; Mismatches 282;   Indels 45; Gaps 5;

Qy   390 rgargararathwsnaarcarcaarwsnathaargargtnacntggtgnytynytnaargcntt 449
          :|:||:||| ||::|:| |: ||| |:| :|:|||||||| | :|| | :||||| ||||
Db   1338 AGGAGAAATTTCTAAGCACACAAAAGCATTCAAGAGAGTGACTTGGGTGCTGTTAAAGACATT 1279

Qy   450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncaycagayaayatnat 509
          :|||||:||| :| :| ||||| :||||| || |||||||||:| | | |:||||||
Db   1278 CAGTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAAATTTGCACCCTGACAAATGAT 1219

Qy   510 haaraaaaraa--yccntytywsngargngnaarttyaarytngcngnagarathgyath 567
          | ||||||||| :||| :| |:| ::| ||||||||:||:|:||||||||| |||
Db   1218 AGAAAGAGAAAAATTCCATTTTCTGAGGAGAAATTCCAAGGTGGCTGCAGAAATTTGCATG 1159

Qy   568 tgyaaygyargaryntnaaygtnaaygccncarcargayaayggngaraayathwsntggnactgy 627
          :|:|| |:| || | || | ||| || ||:||:|| | |||| || |||||||:::||||
Db   1158 AGTAAC-AGGAGCAAAATGCTAAATTCCCAAGACAATGGGGAAAAATGTCTCCAGGCCATGT 1100

Qy   628 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggttc 686
          :|: ||| : || || || |||| |:|| |:||| ||| :| :| ||||
Db   1099 CAGAGCGTCTTTATTGGCAACCCCTCCCATCACAGGTGCCAGAGGTATCAGGAAAAAATGGTT 1040

Qy   687 ytgyggnacngcngnwsnytntgtysygtncarccnmgngayytngtncccntgygt 746
          ||:|:| ||||| :| :| ||:|:| | :||| ||| | ::| |||:|||| |:|
Db   1039 TTGTTGGCCAGGCCCGGGTCCTCATGCTGTGTGCACGGCCTAGGGACTTGGTGCCCCTGCAT 980

Qy   747 nccngtna---------------aywsngcngtngcnwsng 772
          | |:||| |:|| |||| |||:|
Db   979 CCCACCACTCCCACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCCTGTAGCTTCCG 920

Qy   773 arggngacnwsnccnaarccnctggcaryytnccnwsnggngtngarccngtnggngcnaara 832
          |:||:||| :|:|| ||||:|||:| |||| |||:| |||:|||| |: |:| ||:| |||
Db   919 GGAGGTCAAGCCCCAAGCCTTGACAGCCTTCCATGTGTTGGTTGAGACTGCCAGAGTGCACAGA 860

Qy   833 arwsnmgnathgargtntgggaarccnaccnathmgnttycaraacaathtayggnaayccnt 892
          |:|:| ||:| |:||| ||||| ||| |||| :| ||||| || | || |||| |||
Db   859 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGAATGTGCCGAAAATGCCT 800

Qy   893 ggatgccnmgncaararttygcngtnaggngtnggnwsnsntggngnacnwsngcnmgng 952
          ||:||||| :|:| |:|||||:|| ||: ||||| |||:||:| | || :||:| :|
Db   799 GGATCCCAGTCAGAAGAGTTTGCTCAGGAGACGCAGGCCCTCATGGAGATCCTCTGCAGGG 740

Qy   1013 snwsnmgncngtnmgmgnwsncncnccnwsnsmgnytncaraarggngmgnwsnacng 1072
          |||| ||||:|| |:| || :||:|||| ::| ||||| ||||:|||| |||:|||||
Db   739 CAGTGCAGAAGGGAAAATGTGGGGTCAGAGACCCCACACAGTCCTTACTGGGGACCAC 680

Qy   1073 aywsnsnytncarcaygtnccngaraarwsnacngayaacncartgycarccncngtnaargcng 1132
          |:|| :||||||||||||:|:| ||||:|: |||||| ||||| ||:|| ||:| :|||||| :|
Db   679 CTAGTGGGAGCTGTGAGAAGAGGTCCTCC------AGACCCCAGAATGGTAGAATCCACCG 627

Qy   1133 cnggngatggaarwsngtnccntayaaracngtngtngcngaryytnacnacnagcagtngngna 1192
          | :| ||:|||| ||||| ||:|:||| ::|||: |||| | |||| ||:|||||:|:||: ||
Db   626 ACAGCGTTCCACCGTGTCACCTTGACAGACACTCAATCAGGGCCCATAGTCGAATGCAGTTGGGA 567

Qy   1193 thtayytnytncaytgycaygayaytngayytngmgncaygngtgnaarmgnagncaycayttg 1252
          :||: ||||:||| :|:|| ||| | :| ||||:|| |||| || |:|| |:| |||||||:|
Db   566 CCCACCCCTTCCATCCGTGGATGATACGGAGGTCAAAAGGCAGGTCAAAAGGAGAATCATTTTG 506

Qy   1253 gngcnytnmgnttyggaytgycgnacngngnttymgnaactayatgggngcncngtnccnnytnt 1312
          |: :||||:|| ||| :|||::|| :|||||:| | :|||||||: |:| |||:||:|||: :|
Db   505 CCCACCCCTTCCATCCGTGGATGATACGGAGGTCAAAAGGCAGGTCAAAAGGAGATCATTTTG 447

Qy   1313 gyttyggncaarttytyccnttyacncagncngtnttyacncaartgyytntayytncayt 1372
          |:||| ||||| :||::||:||| ||:|| || ||:| || |||||::| |||:|||| |
Db   446 GACCTTTAAGATTTGACTGCCCCACTCGATTTGACTCTCATGGGCCCTCTTT 387

Qy   1313 gyttyggncaarttytyccnttyacncagncngtnttyacncaartgyytntayytncayt 1372
          |:|||:||||| :||:||:||| ||||| || ||:| || |||||::| |||:|||| |
Db   386 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTGATTTACCCAATGCCTGTACCCCATT 327

Qy   1373 gyat 1376
          |:||
Db   326 GTAT 323

RESULT 3
AAS03689/c
ID   AAS03689 standard; DNA; 9236 BP.
XX
AC   AAS03689;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Rhesus gene locus: upstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD;
KW   RHCE; Rh negative; Rh positive; blood group typing; blood transfusion; antigen
KW   C; haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999; 99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA, Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 9; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   upstream Rhesus box of D positives. The Rhesus genes locus

CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RhD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;

Query Match   32.88%; Score 452; DB 22; Length 9236;
Best Local Similarity   47.4%;   Pred. No. 2.7e-103;
Matches   485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

```
Qy  390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtvytnytnaargcntt 449
        ::|||  |: ||||:|| |||::|| ||||||::: ||:|| | |||  |||::|
Db 6014 AGGAGAGAAATTCTAAGCAGCAAAGCAATTCAAGAGAGGTGAACTTGGGTCCTGTGTTAAAGACATT 5955

Qy  450 ywsnttyathmgngargcngarcayaarwsnwswsngaraaaytncaycccnayaaygtnat 509
        :||| |||| |||||||||||||| ||||:|| || ||||||||:||||||||:||
Db 5954 CAGTTTTTATAAGGAGACCAGACATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy  510 haararaaraaa--yccnttywsngargnaarttyaarytngcngngarathtgyath 567
        : ||||| ||||  ||| ||||:||:|||||||||| ||:|:|:|:| ||||:|:|
Db 5894 AGAAAAAGAAAAATTCCCATTTTCTGAGGAGAAAATTCAAGCTGGCTGGCCAGAAATTTGCATG 5835

Qy  568 tgyaaygargarytnaaygtnaayccncaargayaayggngaraayatwsntggactgy 627
        :|: || ||||| |||| |||||:||||:|||| || ||||||||: |||| ||| ||
Db 5834 AGTAAAC-AGGAGCCAAAATGCTAAATTCCCAAGACAATGGGAAAAATGTTCTCCAGGGGCATGT 5776

Qy  628 carmgnwsnqnwsncaarcarwsnathaarwsnytng-cntgmgnqccnmgnmgnaaartggtt 686
        ||| || :| |:||||| |||:|||||| |||:||:  :|:|||| |||||| ||||||||||
Db 5775 CAGAGGTCTTTATTGGCAAACCCCTCCCATNCACAGGTCCTACCAGAGGTGTATCCAGAAAATTGGTT 5716

Qy  687 ytgyggnacnggnccnggnwsnytntcgyttgygtvtncarccnmngnayyytngtnacntgygt 746
        :|| ||| || || ||:||:|| || ||||| ||:|:|||| || || |:|||| |||:||:|:
Db 5715 TTTGTTGGCCAGCCCCGGGGTCCTCATGCTGTGTGCACGGGGACTTGGGTCCCCTGCAT 5656

Qy  747 nccngtna------------aywsngcngtngcnwsng 772
                                :|| :|||| |||||| |:|:
Db 5655 CCCAGCCACTCCCAACCATGACTGACGGGGAGGCAAGGTAGAGAGCCTTGGGCCTGTAGCTTCGG 5596

Qy  773 arggngcnwsncnaarccntggcaryytnccnwsnggngtngarccncngtnggcngncaaara 832
        | || || :| |||||| |||||| :|||||:|| |||:||:| || ||:|||| ||:||||
Db 5595 GGAGTGCAAGCCCCAACCTTGACACGCTTCCATGTGGTGTTGTGCGACTGCAGA 5536

Qy  833 arwsnmgnathgargtnttggarcncncnathmgnttycaraaraarathtayggnaaycct 892
        | |||||||||||||||||||| |||||||||||||||||||||||| | ||||| |||||
Db 5535 AGTCAAGAAATTGGGGTTTGGAAACCTTCGCTTAGATTAAAGAGGAIGTGCGGAAAATGCCT 5476

Qy  893 ggatgccnmgncaraarttygcngtnggngtnggnwsnsntggmgncacnwsngcnnmgng 952
        ||||| || |||||||:|||| || || |||:|:|| |||:|||| |||| ||
Db 5475 GGATGCCCAGTCAGAAGTTTGCTGCAAGAGATCCCTGCTGCCAGGG 5416

Qy  953 tngtncaaraggnaaygtngggntgggarcncncncaymgngtnccncaymgncnncnw 1012
        |:|| |||||| |||||||||||:||||| || || ||:|| || || ||:||:|:|||
Db 5415 CAGTGCAGAAGGAAAATGTGGGTCAAGAGACCTCCTTACTGGGGCACCAC 5356

Qy 1013 snwsnsmgncngtnmgnnwsnccncncnnsnmgnytncaaraarggnmgnwsnacng 1072
        :| ||:|||||||:|||:|| ||:|:| || |||||||||||||:|| ||:||||
Db 5355 CTAGTGGGAGCGTGTGAAGAAGGTCCTCC-----AGACCCCAGAAATGGTRAGATCCACCG 5303

Qy 1073 aywsnytncacraaygtnccngaraarwsnacngayacncartggycarccncgtnaargcng 1132
        |::||:|||||| |||||||||||||| ||||||||:|||||:|||||| ||:|||||||||
Db 5302 ACAGCTTGCACCGTGTTACCTGGAAAAGCTGCAGACACTCAATGCCCAGCCCATGAAAAGCAG 5243
```

RESULT 4
AAS03690/c
ID   AAS03690 standard; DNA; 9236 BP.
XX
AC   AAS03690;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Rhesus gene locus: downstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999; 99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA, Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 10; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   downstream Rhesus box of D positives. The Rhesus genes locus
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant

```
Qy 1133 cngnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
        |:||||||| || |||||||| |||| ||||||:|:||| |||| ||| |||||| ||
Db 5242 CTGAGAGGGAGGCGTGTACCCTGCAAAAGTGACACAGGGACATGCCAAGACCATGGGAA 5183

Qy 1193 thtaytytnaytcaygtgycaygayavytngaygtmngncaygygngtnaarmngnaycayttyg 1252
        |:||:| || |||: ||:||||||| |||||||:| || |||:|||||||| ||| || ||||
Db 5182 CCCAACCCTTCCAATCTGCCTGGACCTGACCTGGAGAGTGAGATCGAGATCACATTTTG 5123

Qy 1253 gngcnytnmgntcygaytgycnacnggngnttymgnaactayatgggnccngtacnnytnt 1312
        | || ||| ||||:|||||| |||||:|:|||||| |||| |||||| || ||||:|||
Db 5122 GAGCTTYAAGATNTGACTGGCCCCACTGGATTTNGAACTCTCACANGGGAGATCAATTTTG 5063

Qy 1313 gyttgyggncaarttytycnttyggnacngcngnttyacncaarttayatnayyytntayytncayt 1372
        |:||:| ||||||||| |||||||| |||||:|||| ||||||||| ||:|:||||| ||:||||
Db 5062 GTTTTGACCAAATTTATCCCATTTGGAATGGCTGTGATTTTACCCAATGCCTGTGTACCCCATT 5003

Qy 1373 gyat 1376
        |:||
Db 5002 GTAT 4999
```

CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;

Query Match 32.8%; Score 452; DB 22; Length 9236;
Best Local Similarity 47.4%; Pred. No. 2.7e-103;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

```
Qy    390 rgargarathwsnaarcarcarwsnathaargargtnacntgggtgtnytytytnaargcntt 449
          :|::||:||||| |:|:: |||:||:|:||||  ||: | | | || ||::  || | |||
Db   6014 AGGAGAAATTCTAAGCACAAACAATTCAAGACGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncaycngayaaygtnat 509
          || |||  || | |||:|||| ||||:|:||| || || | | ||| ||||:|||||| ||
Db   5954 CAGTTTTATAAGGGAAGCAGACATAAAAGTTCCGAAAATTTGCAGCCTCACAAATGTGAT 5895

Qy    510 haaraaraaraa--yccnttywsngargnaaarttyaaarytngcngncnaarathtgyath 567
          :|:|| ::|  | ||| |||| ||||||| | :||||:||| |  | ||| | || | ||
Db   5894 AGAAAAGAAAAATTCCAATTTTCTGAGGACGAAATTCCCAGAAAATTTGCATG 5835

Qy    568 tgyaaygargaarytnaaygtnaayccncarcargayayathwsntggacntgy 627
          |:| ||: || ||| ||| |:|| ||:| |||| | |||||| ||:||| |:||||
Db   5834 AGTAAC-AGGAGCCAAAATGCTAATTCCCAAGACAATGGGAAAAAGTCTCCAGGCCATGT 5776

Qy    628 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgncccnmgnmgnaartggtt 686
          |::|  | | | ::|  ||||||||::|||| | |:||  ||||| ||||| || |||
Db   5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGGTCCAGGTATCAGGAAAAAATGGTT 5716

Qy    687 ytgyggnacnggngnwsnytntgyttgyygtncaarccnmgngayytytngtncntgygt 746
          |:||||| |||||||||||| |||:| :| | |||:|||||||| ||: | | |||| |||
Db   5715 TTGTTGCGGACGCCCCAGGCCCGGGGTCCTCATGCTGTGTGCCAGGGACTTGGTCCCTGCAT 5656

Qy    747 nccngtna------------------aywsngcngtngcnwsng 772
          | ||||  |                  :|:| ||| | || :||
Db   5655 CCCAGCCACTCCCAACCATGACTGACGGAGGACGCAAGGTAGAGCTTGGGTGCTGTAGCTTCGG 5596

Qy    773 arggngcnwsncnaarccntggcaryytnccnwsnggngtngarccngtnggngcnaara 832
          || ||| || | | || |||| ||| |||:| |||||||||||||| ||| ||| || |
Db   5595 GGAGTGCAAGCCCCAAGCCGTGACAGCTTCCATGTGTTGAGACTGCCGAGTGCACAGA 5536

Qy    833 arwsnmgnnathgargtntgggaarccnccncnathmgnttycaraararathtayggnaaycct 892
          || ||| |  | |||| ||| | | ||| |:|:|  || | ||| |:|||| | || ||| |||
Db   5535 AGTCAAGAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGANGTGCGGAAATGCCT 5476

Qy    893 ggatgccnmgncacnaaarttygcngtngngtnggngwsnwsntggmgnacnwsngcnmgng 952
          :: |||| | || | || |||| | |||| ||||||||||| :||| || | ||| ||:|
Db   5475 GGATGCCCAGTCAGAAGTTTGCTGCGAGGACGCAGGCGCCCTCATGGAGATCCTTCTGCCAGGG 5416

Qy    953 tngtncaraargnnaaygtnggntgggarccncnccncaymgngtnccnwsnggngncnccnw 1012
          |:||:| |||| || | | ||||||||||| | |:| ||:|  |||| ||| |||| ||||
Db   5415 CAGTGCAGAAGAGAAATGTGGGGTCAGAGACCCCACACAGTCCTACTGGGGCCACCAC 5356

Qy   1013 snwsnmgncngtnmgnmgnnwsnsnwsnmgnytncaraaraggngmgnwsnacng 1072
          :| |||| | || | || | | |||| | || |:||| || ||:| || | |||
Db   5355 CTAGTGGAGCTGTGAGAAGGTYGAGAGGTCCTCC-----AGACCCCAGAATGGTAGATCCACCG 5303

Qy   1073 ayywsnytncarcayytncngararwsnacngayacncaartgycarccngtnaargcng 1132
          || || |:|| |||| |||||||:||| ||| | | || ||| ||:|| ||||| ||||
Db   5302 ACAGCTTGACCGGTGTTACTTGGAAAAGCTCCAGACACTCAATGCACCCAAGAAAGCAG 5243

Qy   1133 cnggnatggarwsngtnccntayaaracngtngtncgngarytnacnaaracngtnggna 1192
          |:||||| |||| | | |: || ||: |:||||: || || ||||||||: || |||| |
Db   5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGCCAGAGCTGCCCAAGACCATGGGAA 5183

Qy   1193 thtayytytnaayytgycaygayytngaaytmgncaygtnaarmgmngaycaytty 1252
          || |: || || || |:| | | ||||||:||| || || ||| || ||| ||||:|
Db   5182 CCCACCCTTCACTCGCGTGACCTGGATGTGAGTGCAAGGAGGATCATTTTG 5123

Qy   1253 gngcnytnmgnttygaytgycnacngnttymgnacntayatgggncncngtnccncnytnt 1312
          | |:||| || |::|| || | || |  ||:|| || || ||| | | ||| ||| ||||
Db   5122 GAGCTTTAAGATTTGACTGCCCACTGGATTTTGGACTTCCATGGGCGCTTAGCCTCTT 5063
```

Qy   1313 gyttyggncarttytyccnttcyggnacngcngtntcyacancaartgyyytntayytncayt 1372
          ||||||:|:|| ||||:|::| |||||| ||| |:|:||||:||| | ||::| |||| ||:|
Db   5062 GTTTTGACCAATTTATCCCATTTGGAATGGCGTGTATTTACCCAATGCCTGTACCCCATT 5003

RESULT 5

AAS03688/c
ID AAS03688 standard; DNA; 9241 BP.
XX
AC AAS03688;
XX
DT 29-AUG-2001 (first entry)
XX
DE Rhesus gene locus: hybrid Rhesus box of RHD negatives.
XX
KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS Homo sapiens.
XX
PN WO200132702-A2.
XX
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000WO-EP10745.
XX
PR 02-NOV-1999; 99EP-0121686.
PR 31-MAY-2000; 2000EP-0111696.
XX
PI Flegel WA, Wagner FF;
XX
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PS Disclosure; Fig 8; 135pp; English.
XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX
PR WPI; 2001-291052/30.
XX
DR WPI; 2001-291052/30.
XX
CC The sequence represents the coding sequence of Rhesus gene locus:
CC hybrid Rhesus box of RHD negatives. The Rhesus genes locus
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RHD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RHD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other;

Query Match 32.8%; Score 452; DB 22; Length 9241;
Best Local Similarity 47.4%; Pred. No. 2.7e-103;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

```
Qy    390 rgargarathwsnaarcarcarwsnathaargargtnacntgggtgtnytytytnaargcntt 449
          :|::||:||||| |:|:: |||:||:|:||||  ||: | | | || ||::  || | |||
```

```
Db  6014  AGGAGAAATTTCTRAGCAGCAGAAAGCAITCAAGAGGTGACTTGGGTTGCTGTTAAAGACATT  5955
Qy   450  ywsnttyathmgngarrgncngarcayaarwsnwsngaraayytncayccngayaaygtnat   509
Db  5954  CAGTTTTATNAGGGAAGCAGAGCAITAAAAGTTCGGAAAAATTTGCAGCCTGACAATGTGAT  5895
Qy   510  haaraararaa--yccnttywsngargnaarttyaarytngncngarathtgyath        567
Db  5894  AGAAAAGAAAAATTCCCAITTTCTGAGGAGAAAITCAGCTGGCTCGCAGAAAAITTGCATG   5835
Qy   568  tgyaaygyargarytnaaayccncacgayaaygyghgaraayathwsntggacntgy       627
Db  5834  AGTAAC-AGGAGCCAAAATGCTAAITTCCCAAGACAATAGGGAAAAATGTCTCCAGGCCATGT 5776
Qy   628  carmgqwsnqwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnnaartggtt  686
Db  5775  CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCTCAGAGGTATCAGGAAAAAATGGTT  5716
Qy   687  ytgtgygnacnggnccngmnwsnytntgytgyygtncarccnmgnmgayytngtnccntgygt  746
Db  5715  TTGTTGGCCAGGACCCCGGGTCCTCATGCGTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656
Qy   747  nccngtna-----------------aywsngcngtngcnwsng                     772
Db  5655  CCCAGGCCACTCCCAACCATGACTGACGGGAGGGAAGGTAGAGCTTGGGCCTGTAGCTTCGG 5596
Qy   773  arggngacnqwsncnaarccntggcaarytnccnwsngngntngarccngtngggngcnaara 832
Db  5595  GGAGTCAAGCCCAAGGCCTTGACACTTCCATCGTTGGTTGAGACTGCAGTGCACAGA      5536
Qy   833  arwsnmgnathgargtntgggarccnccnathmgntcaraarathtayggnaayccnt      892
Db  5535  AGTCAAGAAATTGGGGTTTGGAAAACCTTCGCCTAGATTAAAGAGGATNGTGCGGAAATGCCT 5476
Qy   893  ggatgcncmgncaararttygcngtngggtngghgnwsnwsntggmgnacnwsngcnmgng   952
Db  5475  GGAATGCCCAGTNCAGAAGTTTCCTGAGGAGGCAAGGTAGAGCTTGGGCCTGTAGCTTCGG  5416
Qy   953  tngtncaaraarggnnaaygtnggntngggarccnccncaymgnmgtnccnwsngngnccnw  1012
Db  5415  CAGTGCAGAAGGGAAAATGTGGGTCAGAGACCCCACAACAGTNCCCTACTCGGGCCACCAC  5356
Qy  1013  snwsnmgncngtnmgqmgnnwsncncncnwsnsnmgnytncaraargngnmgnwsnacng   1072
Db  5355  CTAGTCGGGAGCTGTGAGAAGAGGTCCTCC------AGACCCCAGGAATGGTGAGATCCACCG 5303
Qy  1073  aywsnytncarcaygtnccngaraatwsnacngayacncartgycarccngtgaargcng    1132
Db  5302  ACAGCTTGCACCGTGTACCTGGACACTCAAGTGCCAGCCCATGAAAGCAG             5243
Qy  1133  cnggnatggarwsngtnccntayaaracngtngtcngarytnacnacaacngtnggna     1192
Db  5242  CTGAGAGGGGAGGCTGTACCCTGCAAAAGGTACAGGGCGACGACCACCATGGGAA        5183
Qy  1193  thtayytnytncaycytgycaygaytnmgncaygghgtnaarmgngaycaytygy        1252
Db  5182  CCCACCCCTTCCATCTGCGGTGACCTGGATGTGGAGATGGGAGGAGGATCATTTG       5123
Qy  1253  gngcnytnmgntygaytgycnacngngnttymgnacntayatggggnccngtnccnytnt   1312
Db  5122  GAGCTTTAAGATTTTGACTGCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTTCTTT  5063
Qy  1313  gyttygngncarttyytyccnttygnacncgngtntttyacncaartgytytntaayytncayt 1372
Db  5062  GTTTTGACCAATTTATCCCAATTGGAATGGGCTGTATTTACCAATGCCTGTACCCCCATT  5003
Qy  1373  gyat                                                          1376
Db  5002  GTAT                                                          4999
```

ID   AAD05134 standard; cDNA; 3203 BP.
XX
AC   AAD05134;
XX
DT   17-JUL-2001 (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:24.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; schizophrenia; asthma;
KW   Parkinson's disease; cognitive disorder; Alzheimer's disease;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   153..356
FT   /*tag= a
FT   /product= "Human secreted protein"
FT   sig_peptide   153..215
FT   /*tag= b
FT   mat_peptide   216..353
FT   /*tag= c
FT   /product= "Human mature secreted protein"
XX
PN   WO20013769-A2.
XX
PD   17-MAY-2001.
XX
PF   01-NOV-2000; 2000WO-US30040.
XX
PR   05-NOV-1999; 99US-0163580.
PR   30-JUN-2000; 2000US-0215130.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;
XX
DR   WPI; 2001-308781/32.
DR   P-PSDB; AAE01245.
XX
PT   New isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition -
XX
PS   Claim 1; Page 408-409; 519pp; English.
XX
CC   AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC   protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC   AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   24 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., schizophrenia, asthma,
CC   Parkinson's disease), cognitive disorders (e.g., Alzheimer's disease),
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to

```
CC  sunburn, to maintain organs before transplantation, for supporting cell
CC  culture of primary tissues, to regenerate tissues, to identify their
CC  cognate ligands or binding partners, and in chemotaxis, and can be used
CC  as a food additive or preservative to modify storage properties.
CC  Antibodies specific for a protein of the invention can be used in
CC  alleviating symptoms associated with the disorders mentioned above, and
CC  in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC  immunosorbent assay (ELISA). The present sequence represents a human
CC  secreted protein-encoding cDNA of the invention.
XX
SQ  Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;

Query Match             31.2%;  Score 429.8;  DB 22;  Length 3203;
Best Local Similarity   46.5%;  Pred. No. 4.8e-98;
Matches  476;  Conservative 211;  Mismatches 297;  Indels  40;  Gaps  6;

Qy    391 gargarathwsnarcarwsnathcargargtnacntgggtnytnytnaargcntty 450
          |::|::|||::|:|:||::|::| |::|:|||||||||::|||||| | ||||:|:
Db   1639 gaagaaattctaagcagcaaagcattcaaggagggtgattggggtgctgttaaaagcatt 1698

Qy    451 wsnttyathmgngargcngaarcayaarwsnswsngarayttncaycyngayaaygtnath 510
          |::|:|:||::| |:|||:||::|||||:|::||||||::|:|::||:||:|:|||:||||
Db   1699 cattttaaaaggggaaac--agcataaaaattgggaaattgcagccagttgatgcagca 1756

Qy    511 aaraaaraaraaycnttywsngargngnaarttyaarytngcngngarathtgyathtgy 570
          ::|:| :||:|| ||||||||::|||:|:||:||||||||||:|||||::|:|:||:|:|
Db   1757 gaagagagaaaactgttttttgaggagaaattcaaagctggctgagaaatttgcataagt 1816

Qy    571 aaygargaarytnaaygtnaaycnncoargayayayggnaaraayathwsntggacntgycar 630
          ||:| :| ::|:: |:|||||:|||||::||||||||||||||| :| |||||||||::|:
Db   1817 tacagggaagctgaatgttaatctccaagacaatggggaaaatgtctccaggcatgtcac 1876

Qy    631 mgnwswsnwsncarcarwsnathaarwsnytncntggmgnccnnmmgnaar---tggtty 687
          :::| |:::||::| |: | :|| |||||||||||||||::||:|||:::::|||:|||:|
Db   1877 aggtcttcatggcagccccctccccatcacagaccctgaagactaggagaaaaatggtt 1936

Qy    688 tgyggnacngggnacnggngwsnytntgytgygtncarccmngagyytngtncntgygtn 747
          |::|| |::|::||:|||:||::|::||||||||| |: ||||:||:|:||||:|| || |
Db   1937 tgtgggcaggccaaggtcccagctgctgtgtcagcctaggacttggtgtgcctgcatc 1996

Qy    748 ccngtnaaywsngcngtngc------------------------nwsngarg 775
          |||||:||::|:||||||||                        |:|||||:
Db   1997 ccagctgtcccagccatgctaaaaggcgaggtacagctctgcccatggttcagagg 2056

Qy    776 gngcnwsnccnaarcccntgggcaarytnccnwsngngngtngngncngngcnaaraarw 835
          ||:||:|::|::|||||||:|| ::|||:|||||:|||:|:|:|::|:|::|||::||:
Db   2057 gtgcaagccccaaaccttggcagcttccatgtggtgttggcctgcaaggtgcatggaagt 2116

Qy    836 snmgnnathgaargtntngarcncnccnnathmgnttycaaraarathtaaygngnaayccntgga 895
          :::||||||:||| |:|:|:||::||:| | | :|:|:||:|||:||:|:|||||||||||
Db   2117 caagaatgagggtttgggaacctccatctagattcagaagatgtatgaaatcacctgga 2176

Qy    896 tgccnmgncaaraarttygcngtnggngtngngnwsnwstggmgnacnwsngcnmgnytng 955
          :| ||::|::|::||||:|||:|||:|::||:::||:|::|||||:||:||:|||:|::||
Db   2177 tgctcaggcaaaagttgctccaagggcaagacctctcatggagaacctctgctaggcag 2236

Qy    956 tncaraarggnaaygtngghtngntggarccncncncaymgngtnccnwsngngcncnwsnw 1015
          :|||| |::|||| |:|| |:||::||||::||::||:|||::||:|:|:::||:|||::||
Db   2237 tg-tgaaggaaaatgtggggttggagccccacacagaatccctactgggcaccacta 2295

Qy   1016 smngngcngtnmgnmgnwsnccncnwsnsnmgnytncaraarggnmgnwsnacngayw 1075
          ::::||:||:::|:::|||||:||||||:|:|||:|||:||:|||::|::||||:|||
Db   2296 gtggagctgtcgaagaggccacgtcctcaggacccagatggtagatccaccgaca 2355

Qy   1076 snytncarcaygtnccngararwsnacngayancaartgycatccgngtnaaargcng-cn 1134
          :::|::|::||||:|::||::||::|:|:|:||:|:|:||||::::||:|:|||::|:||
Db   2356 gctctgtaccgtgcacctggaagaagccaagccgtgaaagcaagcca 2415

Qy   1135 ggnatggaarwsncngtncnctayaaracngtngtncngayrytnacnaaracngtnggnath 1194
          :|:|| |||:|::|||:|:::|||::|||||::|:::|| | | :|:|::||:|||||:|
Db   2416 ggagtggggggactatacccctgtgaagccacacaggagctgccaagactaaggaaac 2475

Qy   1195 tayytnytgcaygtgcaygayytngayytngngncaygggngtnaaargngayacytyggn 1254
          ||:||:||::|::|||||:|:||:|||||::||||:|:||||:||||:||:|:||||::||
Db   2476 tacctcctgcatcacatgtgacctgaagacgtgagacatggagtcagagagatcaattttggaa 2535

Qy   1255 gcngytnmgntttygaytgyccnacncggntttymgnacntayatgggnccngt-nccnytntg 1313
          |::| ||::|||:|| ||:|::|:|||:|::| :||||::||||||||||||:| |||:||:|
Db   2536 acgntataaatttgacgcctgctggatttgacctggagcatgggctgaacgctttg 2595

Qy   1314 yttygncaarttytttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
          :|:|:||::||::||:::||::|| |||:||::||:::|::|::|||:|:||:|:|||:||
Db   2596 ttgtgggcagttctcccatttggaatggctgtattacccattacctgtatccccatg 2655

Qy   1374 yatg 1377
          :||:
Db   2656 tatg 2659
```

```
RESULT  7
AAL03635
ID   AAL03635 standard; DNA; 6063 BP.
XX
AC   AAL03635;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6323.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000;  2000US-0179065.
PR   04-FEB-2000;  2000US-0180628.
PR   24-FEB-2000;  2000US-0184664.
PR   02-MAR-2000;  2000US-0186350.
PR   16-MAR-2000;  2000US-0189874.
PR   17-MAR-2000;  2000US-0190076.
PR   18-APR-2000;  2000US-0198123.
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
PR   30-JUN-2000;  2000US-0215135.
PR   07-JUL-2000;  2000US-0216647.
PR   07-JUL-2000;  2000US-0216880.
PR   11-JUL-2000;  2000US-0217487.
PR   11-JUL-2000;  2000US-0217496.
PR   14-JUL-2000;  2000US-0218290.
PR   26-JUL-2000;  2000US-0220963.
PR   26-JUL-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   14-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
PR   14-AUG-2000;  2000US-0225757.
PR   14-AUG-2000;  2000US-0225758.
PR   14-AUG-2000;  2000US-0225759.
PR   18-AUG-2000;  2000US-0226279.
PR   22-AUG-2000;  2000US-0226681.
PR   22-AUG-2000;  2000US-0226868.
PR   22-AUG-2000;  2000US-0227182.
PR   23-AUG-2000;  2000US-0227009.
```

```
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   17-NOV-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   17-NOV-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   14-SEP-2000;   2000US-0234274.
PR   21-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   25-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.

PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   200lUS-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition -
XX
PS   Disclosure; SEQ ID NO 6323; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 6063 BP; 1655 A; 1162 C; 1259 G; 1987 T; 0 other;
```

```
Query Match            31.1%;  Score 427.6;  DB 22;  Length 6063;
Best Local Similarity  46.4%;
Matches  475;  Conservative  211;  Mismatches  298;  Indels  40;  Gaps  6;
                                   Pred. No. 2.9e-97;

Qy    391 gargaratahwsnaaarcarcarwsnathcarargtnacntgggtnytnytnaaargcntty  450
          | :|:||:|||:::|| ::||:|| :||||||:|   || || :|||||:||
Db   3481 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggttgctgttaaagacatt  3540

Qy    451 wsnttyathmgnargcngarcayaarwsnwsngaraayytncayccngayaaygtnath  510
          :||:|||||||| |:|::|::| |:|||:|||:::| |:|:|||:| :||||:||||:
Db   3541 cattttaaagggaaac--agcataaaaatttggaaaaattgcagccagtgatgcagca  3598

Qy    511 aaraaraaraayccnttywsnagrggnaarttyaarytngcngcngarathtgyathgy  570
          | :|:||||:| ||| ||||:|| ||:||:|||:|:|| |:||:||:|||:|:||:| :
Db   3599 gaagagagaaacctgtttttgaggaggaaatccaagctggctgcagaaattgcataagt  3658

Qy    571 aaygargarytnaaygtnaaayccncargacayaayggngaraayathwsntgacntycar  630
          | : :||: :||| :|| :|: |||:|| ||||: || :|||:|| | |:|||:|:|
Db   3659 tacaggggagctgaatgttaatctccaagacacatggggaaaatgtctccaggacatgcac  3718

Qy    631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty  687
          :|:|::  :| || ::| |:  :|:|:||:       ||||:
Db   3719 aggtcttcatggcagcccctccatcacacagacccctgaagactaggaggagaaaatggtt  3778
```

```
Qy    688  tgyggnacnggnccnggnwsnytatgtgyytgyytnccntgygtn                747
           ||:|:|||:|||:|||||||:|:||||:||||::||||:|:|||
Db   3779  tgtgggccaggccaggctcccacatgtgtgtcagcctagggactggtgtgtcagcctaggg   3838

Qy    748  ccngtnaaywsng------------------                               775
           |||: :|                        cngtngcnwsngarg
Db   3839  ccagctgctccagccattgctaaaaggcggccgaggtacagctctgccggttcaagagg   3898

Qy    776  gngcnwsncnaarccntggcarytnccnwsnsggngtnggngcnaaraarw           835
           |:| :||| ||| ::|:|||:|||:|:|:|||||:|||||:|||||||:|||:
Db   3899  gtgcaagccccaaaacctggcagcttccatgtggtcgttgtgtcagggaagt         3958

Qy    836  snmgnnathgargtngaarcncccnathmgnttycaaraaarathtayggnaaycycntgga  895
           :||::||:||||:||||| ::|::|||:||:||:||::|||::||||:|||||:|:|||||
Db   3959  caagaattgaggttgggaacctccatctagattcagaagatgtatggaatcaactgga   4018

Qy    896  tgccnmgncaaraarttygcngtnggngtnaggngnacnwsngcnmngngtng         955
           ||| :||:|||::||||||| :| ||||:|||| |:|||::|| ||:|:|||
Db   4019  tgctcaagccaaagtttgctcatgggagagcacttcatcggagaaccctctgctaggcag  4078

Qy    956  tncararggnaaytgnggntgggarcncncncaymgngtnccnwsngngcncnwsnw      1015
           || ||||||:||||| |||||||| |:|::| ::|||||:|| ||:| :||:|||:|
Db   4079  tg-tgaaggaaatgtgggnttggaagcccccaacagaatcctcactgggcaccacta    4137

Qy   1016  snmgngcngtnmgnmgnwsncnccnwswsnmgnytncaraargngnwsnacngayw      1075
           :|:||:|||||:|||||||||| |::|:::||:|:||||||||:||:||:||:|||:
Db   4138  gtggagctgtgcaagaaggccaccgtcctccagacnaatcggtagatccaccgaca    4197

Qy   1076  snytncarcaygtnccngararwsnacngayacncaartgycarccncngtnaargcng--cn  1134
           | :||||:|| |||||||||||:|||:|||| ||||:||:|| ||:|||||||||:| |:|:
Db   4198  gcttgctaccgtgcacttgacctgaaagccaaagccaagccgtgaaagcagcca        4257

Qy   1135  ggnatggaarwsngtnccntayaaracngtngtngcngarytnacncaaracngtnggnath  1194
           ||:|||||| |||:|||| |||||||||||:|||| |:||||:|| ||:|||||||:|||::
Db   4258  ggagtggggactataccctgtaagccacagggggcagagctgcccaaggggaac        4317

Qy   1195  tayytnytncaytgycayagayytngmgncaygngtnaarmgngaaycaytttyggn      1254
           ||:|::|:|||||:|||| ||:||:|:|:|||||:||||:|||||||||:||||||:
Db   4318  tacctcctgcatcattgtgacctggatgtgagtcaggaggatcattctgga         4377

Qy   1255  gcnytnmgntygaytgycynacngngnttymgnaecntayatggnccngt-nccnytntg   1313
           ||:|||:||||:||||||: ||:|:|:||:|||||||| |||||:|||:| ||:||:||
Db   4378  acgttataattgactgctcgctggattttggacttgcatgggggttgtaacgccttg     4437

Qy   1314  yttyggncarttytyccnttyggnacncgngtnttyacncaartgyytyntatayytncaytg  1373
           |||:|||||:|||||:|||||||:||:|:||||||:|||||:||:||:||:|||:|||:|||:
Db   4438  tgtgtgggcacttctcccattggaatggctgtctgtattaccccattaacctgtatcccattg  4497

Qy   1374  yatg   1377
           :|||
Db   4498  tatg   4501
```

```
RESULT   8
ID   AAS44505 standard; DNA; 22509 BP.
XX
AC   AAS44505;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human LEKTI DNA clone CIT978SKB_94F21 contig 11, SPINK5 exons 1-4.
XX
KW   dermatological; PCR primer; sequencing primer; gene therapy.
KW   eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KW   serine protease inhibitor; atopic disease; Netherton's syndrome; ds;
DE   Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI;
XX
OS   Homo sapiens.
XX
PN   WO200164747-A1.
XX
PD   07-SEP-2001.
```

```
XX
PF   02-MAR-2001; 2001WO-GB00897.
XX
PR   02-MAR-2000; 2000GB-0005098.
PR   03-MAR-2000; 2000GB-0005229.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
XX
DR   WPI; 2001-582149/65.
XX
PT   Determining susceptibility to atopic disease or carrier status of
PT   Netherton's syndrome in humans by identifying variants of or mutations
PT   in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT   inhibitor -
XX
PS   Disclosure; Page 88-94; 123pp; English.
XX
CC   Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC   fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC   SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC   (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC   an atopic disease in a human subject can be detected by screening the
CC   genome for one or more polymorphic variants of SPINK5 gene and/or
CC   expression of a variant LEKTI protein in a tissue. Carrier status of a
CC   subject or development of Netherton's syndrome is diagnosed by screening
CC   for the presence of loss-of-function mutations in the SPINK5 gene. An
CC   expression vector comprising a nucleic acid encoding a serine protease
CC   inhibitor or its functional fragment can be used in screening for
CC   compounds with potential pharmacological activity by determining the
CC   serine protease activity of a protein previously identified as a ligand
CC   of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC   asthma, eczema and hayfever.
XX
SQ   Sequence 22509 BP; 6725 A; 4291 C; 4284 G; 7209 T; 0 other;
```

```
Query Match         31.0%;  Score 427.2;  DB 22;  Length 22509;
Best Local Similarity   47.1%;
Matches  472;  Conservative  207;  Mismatches  303;  Indels  21;  Gaps  7;
Pred. No. 1 1e-96;
```

```
Qy    390  rgargarathwsnaarcaarcaarwsnathcargargtnacntgggtnytytaaargcntt   449
           |:|||||:|:||:|||||||||||||:|||||||||||:||||||||::||:||||||:||
Db  14187  agaagaaattctaagcaagcagcagcattcaagagagatcaactggggtgctcttaaagcgcatt 14246

Qy    450  ywsnttyathmgngargcngarcaayaarwsnwsngaraaytncaycccaccgayaaaygtnat   509
           :|:|:||:||:||||| |||||||:||||:||:|||||||:||||:  | ||:|||||::|||
Db  14247  cagcttttaagggaagcagcaacataaaagttttggaaaatttgcagcctgacagtgtgat   14306

Qy    510  haararara--araaycnttywsngargnaarttyaarytngcncngarathtgyath      567
           |||||||| |||||:|:||:|:||||||:||||||||:||:|:|:|||||:||:||||
Db  14307  aaaaaaaaaaagaaaaaatccatttctgaggagaaaattcaagaccagttgcataaattgcata 14366

Qy    568  tgyaaygaargarytnaaygtnaaycncarcagayayggnaaraaayathwsnwstggacntgy  627
           ||:|| ||||||:||:|| ||||||:|:||:||||| ||:|||||:||:|:|||: |||||:|
Db  14367  agtagcaaggagcctgatgttaaatcccaagacaatgg-gaaaatgtcctccaagagatat   14425

Qy    628  carmgnwsnsnacarcarwsnathaarwsnytncntggmngnaar                 684
           ||:||:|::||| |||||:||||:|||||||:||||:|:|| |||                 tgg
Db  14426  cagagacctttgtgtggcaagccctccattacatgccaagaggtttagagggaaaaatgg   14485

Qy    685  ttytgyygnacnggnccnggnwsnytntgtgyygtncaarccnmmnngaarytngtncntgy   744
           ||:|:|:|||||:|||||||:|||:|:|||||::|||||||||:::|||||:||:|||:||:
Db  14486  ttctctaggcaggcttaggtccccctctgttgtcgagcctcgcgttgaagacactcca     14545

Qy    745  gtnccng----------tnaaywsncgngtngcnwsngarggngngcnwsncnaarccntg   794
           ||:||||          |||||:|||:||||||:|||||||||:|:|:|:||||||||:||
Db  14546  gtggctgaaggggggccatgggctatggcttccgagggtgcaagcctgaaagccttg       14605

Qy    795  gcarytnccnwsnsngngtngarccngtnggngcnaaraarwsnmgnnathgargtntggga  854
           |||:|||:|:||:|:|:|||||:||||||||:|:|||||||:||:||::||:||||:||||:
Db  14606  acagcttccacatggtgttgggcctgttgggagtgcacagaagtcaagaactgaggttgaa   14665
```

```
Qy    855  rccnccnathmgnttycararaarthtayygnaaaycntggatgcccmgncararaattygc  914
           :::||:|| ||:|:||:||| ||:.:||::||  ||  ||:||:||||
Db  14666  accctcagcctagatctc---agatatatggaaatgcctgtgatgtccaggcagaagtttc  14722

Qy    915  ngtnggngtnggnwsnsntggmgnacnwsngncnmgngtngtncararaargngnaaygtgg  974
           :|| |:||:|.:|    ||:.:|    ||| |:||::||||:|||:|:|:|:||:||:||||
Db  14723  gtgtaggggtgaggtcctcaggggaacctctgtaggcagtacagacagaagaaagtttc  14782

Qy    975  ntgggarccnccncaymgntng nccnwsngncwsnsnmngncngtnmgnmngnws  1034
           :|::||||||::|:||:||:||:|:||:||:||:||:||:||:||:||:|:||
Db  14783  gtaggagcccccaccacagagtcccctactgg-gcaccactggttgagctgttagaagaga  14841

Qy   1035  ncncccnwsnwsnmgnytncacragrgnmgmgnwsnacngaywsnytncarcaygtnccnga  1094
           ||:||::|   |||:|| || ||::|  || :|||:|:|| ||  :||:|||||||:|:||
Db  14842  gcctttcatcctccagaacccagcaataacagatcccaggttagcttgcacacatatgcctga  14901

Qy   1095  raarwsnacngayacncaartgycacnccngtnaagcgngaatggarwsngtnccnta  1154
           ::|:|| :|| |:|.| ||||| ||:|||||:.:|:||:||:||:|:|||
Db  14902  gaaagccacaaacacttcagtccagcctgtgggaaggggtaccctg  14961

Qy   1155  yaaracngtngtngngarytnacnaaracngtnggnathtayytnaytncaytgycayga  1214
           :||:|.|||||:|||:.||::||||:||.||:||||||:||:||:||:||
Db  14962  caaagccacagaggtggagttgccaagaccaatgggaacccacctcttgatcagcaatga  15021

Qy   1215  yytngaygtnmgncayggngtnaarmgnaycayttyggngncnytnmgnttygaytgycc  1274
           ||:|||:|||| ||:|:||:|||:||:||||:|:||::||||:||||:||:|
Db  15022  cctggataatgaagacatagatccaaaatatatcattttggaacttaagattgactgcct  15081

Qy   1275  nacngngttymgnacantayatgygnccngt-nccnytntgtttyggngncarttytyccnt  1333
           :|:|||::||:||||||:||:|||| |||||  |:|| ||:|:||||:||||:||:|||
Db  15082  tcctgaatttcgacctgcatggcatctgtagcccctttgttttggccaatatctcccat  15141

Qy   1334  tyggnacngcngtnttyacncaartgyytntaytncaytgyat  1376
           :||:|| |||:||||| || |:|||:|:|:|:|::|| ||:||
Db  15142  ttggaacagctgtattgaccaaatgyctgtaccccattgtat  15184
```

```
RESULT   9

ID     AAS44506 standard; DNA; 31529 BP.
XX
AC     AAS44506;
XX
DT     18-DEC-2001 (first entry)
XX
DE     Human LEKTI DNA clone CIY978SKB_94F21 contig 8, SPINK5 exon 5.
XX
KW     Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW     serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW     eczema; hayfever; antiastmatic; antiallergic; antiinflammatory;
KW     dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS     Homo sapiens.
XX
KW
XX
PN     WO20016747-A1.
XX
PD     07-SEP-2001.
XX
PF     02-MAR-2001; 2001WO-GB00897.
XX
PR     02-MAR-2000; 2000GB-0005098.
PR     03-MAR-2000; 2000GB-0005229.
XX
PA     (ISIS-) ISIS INNOVATION LTD.
XX
PI     Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
XX
DR     WPI; 2001-582149/65.
XX
PT     Determining susceptibility to atopic disease or carrier status of
PT     Netherton's syndrome in humans by identifying variants of or mutations
PT     in SPINK5, a gene encoding lympho-epithelial Kazal-type related
```

Right column:

```
PT     inhibitor -
XX
PS     Disclosure; Page 95-103; 123pp; English.
XX
CC     Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC     fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC     SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC     (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC     an atopic disease in a human subject can be detected by screening the
CC     genome for one or more polymorphic variants of SPINK5 gene and/or
CC     expression of a variant LEKTI protein in a tissue. Carrier status of a
CC     subject or development of Netherton's syndrome is diagnosed by screening
CC     for the presence of loss-of-function mutations in the SPINK5 gene. An
CC     expression vector comprising a nucleic acid encoding a SPINK5 protease
CC     inhibitor or its functional fragment can be used in screening for
CC     compounds with potential pharmacological activity by determining the
CC     serine protease activity of a protein previously identified as a ligand
CC     of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC     asthma, eczema and hayfever.
XX
SQ     Sequence, 31529 BP; 9686 A; 5798 C; 5934 G; 10111 T; 0 other;
```

```
Query Match     31.0%;  Score 427.2; DB 22;  Length 31529;
Best Local Similarity  47.1%;  Pred. No. 1.5e-96;
Matches   472;  Conservative  207; Mismatches 303;  Indels  21;  Gaps   7;

Qy    390  rgargarathwsnaaarcacwsnathcaargaytnacntgggtnytnytnaargncntt  449
           :|::|||||:|| :|::||::|||:||:||:|| ||:||:||:|:|||:||
Db  14187  agaagaaatttctaagcaacgaagacattcaagagataactctgggtgctcttaaaggcatt  14246

Qy    450  ywsnttyathmgngarcngarcaryaarwsnwsngataraayytncayccngayaaygtnat  509
           :|::|||| ||:||:||:|| ||:||||:|:||:||||||:||:|:||:|:|||:|||
Db  14247  cagcttcttaaggaaggcaagaacataaaagtttgggaaatttgcagcctgacagtgat  14306

Qy    510  haaraara--araaycnttywsngargngnaarttyaarytngcngcngarathtyayh  567
           |:|||||  :||::|:|||:||:||:||:|||:|:||:||:|||||:||:||:|||
Db  14247  aaaaaagaaaatccatttctgaggagaaattcaagccagttgcataaattgcata  14366

Qy    568  tgyaaygaargaarytnaaygtnaaycncaargayvaaygggnaraayathwsntggacntgy  627
           |:|: |:|||||| |||:|||:||:||:| ||||::|:||||:.:|||||::||:||
Db  14367  agtagcaaggagaccctgatgtttaatcccaagacaatgg-gaaaatgtctccaagagatat  14425

Qy    628  carmgnwswsnncarcarwsnathaarwsnytngcntggmgncnmgnmgnaar---tgg  684
           ||:| |:||:||:|:| |:| |:||| :|:||||::||:||||:|:.||| |||
Db  14426  cagagaccttgtggcagccccccccattacatgccagaggttaggaggaaaaatgg  14485

Qy    685  ttytgyggnacngccnggnwsnytntgytgygtncarccnmgngaytgtgtncntgy  744
           :|::|| ||::|: || ||:|:|||:|:||:|:| ||:|| |:||:||:|:||:|:|
Db  14486  tctctaggccagcttaggtgccctcgtgtgcagcccctgaaagacactcca  14545

Qy    745  gtnccng---------tnaaywsngcngtngcnwsngargngcngcngcnwsnccnaarccntg  794
           |:||:||         |:|| |:||:||:||:||:||:||:||:||:||:||:|:||:|||
Db  14546  gtggctgaaggggggccatggcttgggctatgtcttccgagggtgcaagcctgaaagcctg  14605

Qy    795  gcarytnccnwsngngtngarccngtngngcnaataraarwsnmgnathgargtntggga  854
           ||:|||:||:|:||:||:|| ||:||:|:||:|||:|||:||:||||:||:||:||:|||
Db  14606  acagcttcacactggtgttgagcctgagtgcacagaactgaggtttgaa  14665

Qy    855  rcnccnathmgnttycararaarthtayygnaaaycntggatgcccmgncararaattygc  914
           :|:|::|| || |:|:||:||||| ||| |||||:|:||||| ||:|:||:||:||:|||
Db  14666  acctcagcctagatctc---agatatatggaaatgcctgtgatgtccaggcagaagtttc  14722

Qy    915  ngtnggngtnggnwsnsntggmgnacnwsngncnmgngtngtncararaargngnaaygtgg  974
           :|:||:||:||:||:||:|||:||||:||:||:||:||:||:||:||:||||:||||:||:|
Db  14723  gtgtaggggtgaggtcctcaggggaacctctgtaggcagtacagacagaagaaagtttc  14782

Qy    975  ntgggarccnccncaymgntngnccnwsngncwsnsnmngncngtnmgnmngnws  1034
           :|| ||||||:|:||:|:||:| ::|:|:|:||:|:|:|||||:|:|:|:||:||:
Db  14783  gtaggagcccccaccacagagtcccctactgg-gcaccactggttgagctgttagaagaga  14841

Qy   1035  ncncccnwsnwsnmgnytncaragrgnmgmgnwsnacngaywsnytncarcaygtnccnga  1094
           ||:|||:|||:||:||:| || ||::|  || :|||:|:|| ||  :|| :|||||||:|:
Db  14842  gcctttcatcctccagaacccagcaataacagatcccaggttagcttgcacacatatgcctga  14901
```

```
Db  14842  gccttcatcctccagaaccagaatagacagatccacaggtagcttgcaccatatgcctga  14901
           :||::.||.||||:::||.||:.||||.|||.:.||.||
Qy  1095   raarwsnacngayacncartgycaarccngtnaargcngcnggnatggarwsngtnccnta  1154

Db  14902  gaaagccacaaacactcagtgcaacctctgtgaaggcagctggaggggaggcggtaccctg  14961
           |||::.|||.||:||||:::||:||.|||||.||
Qy  1155   yaaaracngtngtcngaarytnacnaaracngtnggnnathtayytncaytgycayga    1214

Db  14962  caaagccacagaggtgagttgcccagagaccatgggaacccaacctcttgcatcagcatga  15021
           :.||::.|||:.||:|||.:||:::||.|||.|
Qy  1215   yytngaygtnmgncayggngtnaarimgnaycaytyggngcnytnmgnttygaytgycc   1274

Db  15022  cctggattatgagacatagattccaaaaatatatcattctggaacttaagattgactgcct  15081
           ||:||.||:|:.|||.||||.|.||.:.|||:||||.||||||.||.|||.|||||.|
Qy  1275   nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncaarttytycnt   1333

Db  15082  tcctgaatttcgacttcgcatggcatctgtagcccccttgtttttggccaaatcctcccat  15141
           ||||.|||||||.|||:||:||||:.|||:||:|||||||||||.||.||:|||||||||
Qy  1334   tyggnacngcngtnttyacncartgyytntayytncaytgyyat                 1376

Db  15142  tggaaacagctgtattgacccaatggctgtaccccccattgtat                 15184
```

RESULT 10
ID   AAL03634 standard; DNA; 6063 BP.
XX
AC   AAL03634;
XX
DT   21-NOV-2001 (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6322.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228987.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   05-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0233398.
PR   14-SEP-2000; 2000US-0232299.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0233401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   29-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0244474.
PR   08-NOV-2000; 2000US-0244475.
PR   08-NOV-2000; 2000US-0244476.
PR   08-NOV-2000; 2000US-0244477.
PR   08-NOV-2000; 2000US-0244478.
PR   08-NOV-2000; 2000US-0245523.
PR   08-NOV-2000; 2000US-0245524.
PR   08-NOV-2000; 2000US-0245525.
PR   08-NOV-2000; 2000US-0245526.
PR   08-NOV-2000; 2000US-0245527.
PR   08-NOV-2000; 2000US-0245528.
PR   08-NOV-2000; 2000US-0245532.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.

```
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   05-DEC-2000; 2000US-0255719.
PR   06-DEC-2000; 2000US-0251479.
PR   17-NOV-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   17-NOV-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
     is used in preventing, treating or ameliorating a medical condition -
PS   Disclosure; SEQ ID NO 6322; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 6063 BP; 1656 A; 1163 C; 1262 G; 1982 T; 0 other;
```

```
Query Match             31.0%;  Score 426.4;  DB 22;  Length 6063;
Best Local Similarity   46.4%;  Pred. No. 5.9e-97;
Matches  475;  Conservative  210;  Mismatches  299;  Indels  40;  Gaps  6;

Qy        391 gargarathwsnaarcaarwsnathcaargargtnacntggtnytnaargcntty 450
              ||:|||||||:|  |||::||:|:|||:|:|:|| |:|||| |||||:::
Db       3481 gaagaaattctaacagcaaagcaattccaagagagtgattgggtcgtctaaaagcatt 3540

Qy        451 wsnttyathmngcargcngarccayaarwsnwsngaraaytncayccngayaaygtnath 510
              |::|| |:| ||:||||||::|:|  ||||::||||| :|  |::|:|| |:|  :||
Db       3541 catttaaaagtgaaac--agcataaaatttgaaatttgcagccagttgatgtgcaca 3598

Qy        511 aaraararaaayccnttywsngargngnaarttyaarytngcngarathtgyathtgy 570
              || :| ||||||::| |||||:|:|::|||||||| :||||||:|||:| |||:| ||
Db       3599 gaagaagagaaacctgtttttggagagaaattcaagtctggctgcagaaatttgcataagt 3658

Qy        571 aaygaargarytnaayccncarcayaayggngaraaatathwsntggacntgycar 630
              ||:|| ||||:|| |:||:|| ||:|| |||||||| |||:|||:||:||:||:||
Db       3659 tacaggagagctgaatgttaatctccaagacaatgggaaaatgtctccaggcgcatgtcac 3718

Qy        631 mgnwsnwsncarcarwsnathaaarwsnytngcntggmgnccnmgmgnaar---tggtty 687
              ||:||::||:|:||:||:|| |||||:|:||||:||| ||::||:||:| |   ||:||
Db       3719 aggtcttcatggcagcccctcccacatcacagaccctgaagactaggaggaaaacggtt 3778

Qy        688 tgyggnacnggncccnggnwsnytntgytgygtgtncarccnmgnayytngtncntgygtn 747
              ||::|:||:||:|||:||:||:|:|||:||:|:|:|||:||:||:||:|:||:|||||:||
Db       3779 tgtgtgggccaggcccaggtcccactgctgtgtcagcctggttggccctgcatc 3838

Qy        748 ccngtnaaaywsng----------cngtngcnwsngarg 775
              ||:||:|||:|||          ||:||:||:||:||:
Db       3839 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccgtggttcagagg 3898

Qy        776 gngcnwsnccnaarccnntgycaryctnccnwsnggngtngarccngtngggncnaaraarw 835
              |:||:||:||:|||||:|||:|| :|:||:||:||:|:||||:|||:|:||:||:|||||||
Db       3899 gtgcaagccccaaacctttgcagctccatgttcattgttggccttgatatggaaagt 3958

Qy        836 snmngnathgargtntnggarccncnacathmgnttycararathtaygnaaaytncntgga 895
              ||:||:|||||:|:|:|:|||||:|:||||:||:|||||||||:|||||:|||:|:||:|||
Db       3959 caagaattgagtttgggaacctccatctagattcagaaagatgtatggaatcacctgga 4018

Qy        896 tgccnmgncaraarttcygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 955
              ||||:||:|| |||||:|| ||:|||:||:||:||:||:|||||:||:||:|| ||:||:
Db       4019 tgctcaggcaaaagtttgctccaggggcaagaacctctcgctaggcag 4078

Qy        956 tncaraargnaaaytgtnggntgggarccncnacaymgnttycararathtaygnaaytncntga 1015
              |:||||||::|||:||:||:||||||||:|:|||:||:|||||||||:||||||:||:|:||:|
Db       4079 tg--tgaaggaaatgtggggttggagccccacacagaatcccctactgggggcaccacta 4137

Qy       1016 snmgncngtnmgmgnwsncncnwsnwsnmgnytncaraarggnmgnwsnacngayw 1075
              ||:|:|:||:||:||:||:|:||:||:||:|:||||||||||:||:||:|||||:
Db       4138 gtggagctgtgcaaagaagccaccgtgtcttcagacccagaaatggtagatcccacga 4197

Qy       1076 snytncarcaytncngaraarwsnacngayacnacrtgycarccncngtnaargcng-cn 1134
              ||:|:|| ||:|:||||:|||:||:||:|| |:|||:||:|||||:|:||:|||||:| ||
Db       4198 gcttgtaccgtgcacctggaaaaacgcccagcccgtgaaagcagcca 4257

Qy       1135 ggnatggaarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
              ||:|||||||:||:|:||:||:|||:|:||:||:||:||:|:||:|||||:|:||:||:||
Db       4258 ggagtggggactatacctgtgaaggcacagggcagagctgcccaagactaaggggaac 4317

Qy       1195 tayytnaytncaytgycaygayytngmgncayggngtnaarmgngacaytygtn 1254
              ||:|:||:|:|||||:||||| :||:||:||:||:|:|||:|:|:|:||:|||
Db       4318 tacctcttgcatcattgtgacctggatgtgaagacatgggagtcagagagatcattttgga 4377

Qy       1255 gcnytnmgnttygaytgycnacngngnttygnacntayatgggngccngt-nccnytntg 1313
              ||:|:||:||:||:||:|:||:|:|:||:|:||:||:|||||:||:||: |||:|:||:
Db       4378 acgttatataatggactgccgctggattttggactttgctcatggggtgtaaacgccttg 4437

Qy       1314 yttyggncarttytyccnttyggnacngcngntnttyacncaartgyytntcaytncaytg 1373
              :||:||:|||||:|:|||||:||:||:||:|:|:||:||:|||:||:|:||:|:|||:||
Db       4438 ttgtgggcacttctcccattggaatggctgtattacccattaccctgtatcccattg 4497

Qy       1374 yatg 1377
              :|||
Db       4498 tatg 4501
```

```
RESULT 11
AAH13678
ID   AAH13678 standard; cDNA; 1736 BP.
XX
AC   AAH13678;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:10539.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 10539; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide comprising a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primer sets allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;
XX

Query Match             30.6%;  Score 421.4;  DB 22;  Length 1736;
Best Local Similarity   46.2%;  Pred. No. 3.9e-96;
Matches  474;  Conservative  211;  Mismatches  293;  Indels  47;  Gaps   6;

Qy    390 rgargarathwsnaarcarcarwsnathcargargytnacntgggtnytnaargcntt 449
          ::|:|:|| :|:|| :|||: :| || ||||||: || |||||||:| | |:|||| ||
Db    350 gagaccttcaccacagcccctcctatcacagggccaga-----aggaaaagtggtttg 404

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncaycyncgayaayytnat 509
          ::::|||:|| :::|| |:|||:| | :|||| |:|| ||:| |:| :||| | ||:||
Db    405 tgggcctggtccaggtccccctgctgtgtgtgcaacctaggggatggtcgaccctgtccc 464

Qy    510 haaxararaaraayccnttywsngargngaarttyaarytngcngarathtgyathtg 569
          : :|||||:| |::|| :| :| | ||||: ||| |:|: |::|::|:|||| |:|| |
Db    465 agctgctgcttcagctgtggctgaaccatgtacagctcaggttgtggcctcaga 524

Qy    750 ngtnaayws---------------- 773
          |:|
Db    465 agctgctgcttcagctgtggctgaacc---ngcngtngcnwsnga 524

Qy    170 tagtttatataagggaagacagacatacaaaagtttggaaaaatttgcagcattactac 229
          |:|: :|| :|||:|:|||:| ||:|| ||||||:||||:| | ||:|::|| ||
Db    642 agatgccaggcaaaagtttgctat-9999tg9999ccccatcatagaagaacctctgtaagg 700

Qy    893 ggatgccnmgncaraarttygcngtngggntngngwnwsnwsntggmgnacnwsngcnmgng 952
          ||||||| :|::| ||||:|:|| |||||:| |:||: ||: ||:| |:|:|| ::||:
Db    642 agatgccaggcaaaagtttgctat-9999tg9999ccccatcatagaagaacctctgtaagg 700

Qy    110 agaagaaatttctaagcagcacattcaaagaggtgattggatactggttaaggcatt 169
          :|||::|||| :||||:|| | :| ::||::||||:||||||:||| |||||| ||
Db    690 tgggcctggtccaggtccccctgctgtgtgtgcaacctaggggatggtcgaccctgtccc 749
```

PN   WO200162969-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05358.
XX
PR   22-FEB-2000; 2000US-183756P.
PR   20-OCT-2000; 2000US-0692414.
PR   24-JAN-2001; 2001US-0768184.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
XX
DR   WPI; 2002-041152/05.
DR   P-PSDB; AAG68251.
XX
PT   Novel variant of estrogen receptor alpha polypeptide useful for
PT   determining the biological activity of a protein for high throughput
PT   screening and for raising antibodies that elicit an immune response in
PT   host.
XX
PS   Example; Fig 1 page 1-93; 333pp; English.
XX
CC   The present invention describes an isolated peptide (I) consisting of an
CC   amino acid sequence selected from: (a) the amino acid sequence of a
CC   variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251;
CC   or (b) a fragment comprising at least 10 contiguous amino acids of the
CC   protein in AAG68251. (I) has cytostatic, osteopathic, cardiant and
CC   vasotropic activities, and can be used in gene therapy and vaccine
CC   production. (I) is useful for identifying an agent that binds to (I), by
CC   contacting (I) with an agent and assaying the contacted mixture to
CC   determine whether a complex is formed with the agent bound to the
CC   peptide. A polynucleotide (II), encoding (I), is useful in the
CC   development of diagnostics and therapies for diseases and disorders
CC   mediated/modulated by an oestrogen receptor (ER). (II) is also useful in
CC   gene therapy for treating cancer, osteoporosis and cardiovascular
CC   diseases. The human ESR-alpha gene is located on chromosome 6. The
CC   present sequence represents the human ESR-alpha gene, which is given in
CC   the exemplification of the present invention.
XX
SQ   Sequence 465237 BP; 133988 A; 89578 C; 93946 G; 147721 T; 4 other;

Query Match   30.6%; Score 421; DB 24; Length 465237;
Best Local Similarity   46.5%; Pred. No. 5e-94;
Matches 477; Conservative 213; Mismatches 292; Indels 43; Gaps 7;

Qy   392 argararthwsnaaarcarcarwsnathcargargtnacntggtgytnytnaargcnttyw 451
          :!:!!::!!!:::!!!! |!:||| | |!||||:|:::|!::!:!:!::| ! !!::
Db 438788 aagaaatttcaagcaagcaaacaccattcaaaaggtgacttgggtgttgttaaagcattct 438847

Qy   452 snttyathmgngargcngarcayaarwsnwsngaraaaytncayccngayaaygtnatha 511
          ||!|:!!::||:! !!!|!||:||:|:::::||||||||!|::!!||!|:||! | |
Db 438848 gttttaaaagagaaac--agcataaaaagtcagaaaaattcagcgtgatgcagtag 438905

Qy   512 araaraaraaycc-nttywsngargngaarathtgyathtgy 570
          :!!:!:!:!!!::!||!!!!!||||:||:||!||!||:||!:!
Db 438906 gaaagaaaaacccattcttttggaggagaaattcaagctggctgcagaaattgcataagt 438965

Qy   571 aaygargaryrtnaaytnaayccncargayaayathwsntgacntgycar 630
          |::!|!::!::::|!!!| !!::|!:|||:||||||||:!||!::||!
Db 438966 aacaaggagcaaaatgttaatcccaagacaaatgtcccaagagcatgtcat 439025

Qy   631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg----naartggtt 686
          :|:| :||!:|| !||:||!|||||!|!! |||!||!!|||||:| !|! |!|||:||
Db 439026 aggtcttcatgcagccctccacatccacaccgaagcctaggagaaaaacagtt 439085

Qy   687 ytgtyggnacnggngnncgggnwsnaythaayccncargaaytgtycngtcnctgyygt 746
          :!:!| |:| |!!||:|!::| ! !||! || !|::||:|! !| !! |!!
Db 439086 ttgtggggccagtcccaggtcccacctagaacttggtccctgcat 439145

Qy   747 nccngtnaaysngcngtngc---------------------------nwsngargngncnws 782

Qy   1372 tgyat 1376
          ||:||
Db 439805 tgtat 439809

Qy   1312 tgyttyggncarttyttyccnttyggnacncngtnttyacncaartgyytntaayytncay 1371
          |:||!!|!||:||:|!||!!|||!||:!||!||!|||!|!|!:|:|!::!::!!||||:|
Db 439745 tgttttggccaattctcccatttggggctgctgtatttacccaatgcctaaacccgcat 439804

Qy   1253 gngcnytnmgnttygaytgycnacngngnttymgnacntayatgggngccngt-nccnytn 1311
          |!|||:!|!|!||:|!|||!|:|||||!||||||:!|!||!|!|||!!|!| |!|!||!
Db 439626 cctacctcttgacteggagetgtgagacattcaagtcaaagagat-attttg 439684

Qy   1193 thtayytnytncaytgycaygaytgtngncaygtgataarmgngaycaytyg 1252
          |:|:!||!|!||!|:|!|||!||||!||!|||||!|!|||||!|||||!|:|
Db 439566 aggtggtggtgctataccctataaagccacaggggcagagctgccaagctatggaa 439625

Qy   1139 ----------tggaarwsngtnccntayaarcngtngtncngarytnacnaaracngtnggna 1192
          ---------- ||:|!|!|!|:||!|!!|!||:!||:|!||!||:|!!||:|!!|||||
Db 439506 ccatgtgcctgaaagccacacactcaaacgccagcctgaaacgcatcagggttgg 439565

Qy   1083 rcaygtnccncngararwsnsnacngayacncaartgycarccngtnaaargcngcnggna--- 1138
          !|||!||!|!|!!|!!:|||!|!|!!:!|||:!|!:!|::!!:|!||!|:!:||!||||-
Db 439446 tgtaaggagagagggaccactgtccccaacccgcagaaatagtagatccactgacagcttgca 439505

Qy   1023 ngtnmgngmngwsncncncwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnca 1082
          :!:||!!!|||!||!:!:!||!::| !|!::!!||!|||!!|:!||!|:!|:!|!!|||
Db 439386 gggaaaatgtggggttggaacctccaacttatatttcagaagatacgtggaaatgcct 439445

Qy   963 rggnaaygtnggngtggarccncncncaymgngtncccnwsnggngcnccnwsnwsnmgngc 1022
          !||!|!|||:|!|!||||!|!|!||||!|:|||!|:|!||!:||||!||!|!!:|:|!||!|
Db 439326 gcaaacattgctgcagaggtgggggcctcatggaggggccttctgctaggtgccaatgaggaa 439385

Qy   903 ncaraaattygcngtngtngngngtnggmgnnacnwsngcngngttgtgttncaraa 962
          :!:| |!!::!|:|!||!||||!!|||!|||:|:|!|||!!|!!|||||!||!::!
Db 439266 tgagtttgggaacctccaacttatatttcagaagatacgtggaaatgcctggatcaccag 439325

Qy   843 hgargtnttgggarccncnaathmgnttyaraaarathtaygngnaayccntggatgccnmg 902
          :!:||!|!!!!|||!!||!::!!:::|!!!:|!|!!|:!!| |!!|:|!| |||! |
Db 439206 cccaaacccttggcagcgtttcatcgtggttgttgaagcctgtgttgttacacagaagttaagaaat 439265

Qy   783 nccnaarccntggcaryyntnccnwsnggngtngarccngtngggngcnaaraarwsnmgnat 842
          :!|!!| ||!!!| !||:!!| ! !::|!||| |!|!!|!|!!!||!|| :!!:!||!!!:
Db 439146 cctagctgctccagctattgctaaaaggagtgaggtaccacggttcaagaggttgcaag 439205

RESULT 13
AAS91984
ID   AAS91984 standard; cDNA; 1485 BP.
XX
AC   AAS91984;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27788.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.

```
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27797.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID No 27788; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (II) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement.(II) and its binding partners are useful for treating
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note:The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1485 BP; 442 A; 333 C; 373 G; 337 T; 0 other;
```

```
Query Match             29.7%;  Score 409;  DB 23;  Length 1485;
Best Local Similarity   46.5%;
Matches  458;  Conservative  199;  Mismatches  287;  Indels   40;  Gaps    6;

Qy    433 gtnytnaargcnttywsnttyathmngargcngaracayaarwsnswsngaraaaytn 492
          || |||  || :||:||| | ||| ||:|| | ||||:| |||::| ||::| ||||
Db    103 gcttctttgtatgccattcagtttataaggagagcataaaaattggaaattg 162

Qy    493 cayccngayaaygtnaathaaraaraaraayccnttywsngarggnaarttyaarytngcn 552
          |::| | || |:|| ||| | || ||:|||:|| ||:||||| |:|| |||:|| ||:|
Db    163 cagcctgactatgcaatagaaagaaaaccatttctggggagaaactcaagcctatt 222

Qy    553 gcngarathtgyathtgyaayatygaaygarytnaaaaycncnacgayayayggngaraay 612
          ||:| ||:|| ||:|| |||:|| || | | |||||::| |:|| || | ||:| |||||
Db    223 gcagaaattgcctaagtagcaaggagcctaatgttaatcccaagaccatcgggaaat 282

Qy    613 athwsntgacntgycarmgnwswsncarcarwsnathaarwsnytngcntgmgncn 672
          |:|| | || ||:|| |:|:|:|:|  || ||:|| |||:|| ||||||:|:||
Db    283 gtctacaggccatgtcaggaacttcatggctgccctccaatcacaagcctcaggccc 342

Qy    673 mgnmg---naartggttytgygngnacngnccngnwsnytntgytgyygtncaarccnmgn 729
          :|:|:      |:||||:||:|:||||:|||:||:|:| ||:||:|| ||:|| |||:|
Db    343 aggagtaaaagtggttcatggccagcgccaggttcctgtgccgtgcagcacagg 402

Qy    730 gayytngtnccntgyttnccng------------------tnaay 756
          ||:||| || ||:|| |||||                  |:||:
Db    403 gacttggtgccctgttcccagctatggctgaaagggccaagtagaggt 462

Qy    757 wsngcngtngcnwsngarggngcnwsnccnaarccntggaarytnccnwsnggngntngar 816
          |:|||:||:||:|:|| ||:||:|:| ||||||||:||:|| ||:||:|:|||:|:| ||
Db    463 tgggctgtggcttcagaggtgcaagaccaagccttggcagttcaacgtgatgtgag 522

Qy    817 ccngtnggngcnaaraarwsnmgnathgargtntggarccncnacnathmgnttycaraar 876
          |||||:||:||:|||||:|:|:| ||:|||:|||| ||:| ||:|| ||:||:|| |||:
Db    523 cctgtgggtgcacagaagtcaagaactgaggtttggggaactctgcctagatttcagaag 582

Qy    877 athtayggnaayccnttygatgccnmgncaraarttygcngtnggngtnggnwsnwsntgg 936
          |||:||  |||:||:|| ||||||:|:| || |||:||||:|  |:|:|:|| | | |||
Db    583 atgtatggaaacgcctggatgccccaggccaaagttgctgcagctggggaccgtggg 642

Qy    937 mgnaacnwsngcnmgngtngtncaaraarggnaaygtnggngttggnccncnacaymngnt 995
          :|||| | | ||:|:|| |||:|| |||||:||:||:||:|| |:|| ||:| | |:|:|
Db    643 agaacctcctgcctaggcagtgcaggaagaaaatgtgggtcaggaagccccacacagagt 702

Qy    996 nccnwsnggngncnccnwsnwsnmgnggncngtnmgnwsnccncnnwsnwsnmgnnytnca 1055
          |:|||:|||:|:| ||:|:|:|:| ||:|:| :|:|:|| || | |:|:|:| ||:||| |
Db    703 cctactgggcaccactagtggagctgtgaagggggccactgtccctccagaccca 762

Qy   1056 raarggnmgnwsna--cngaaywsnytncarcaygtncngararaarwsnacngaycncar 1114
          | || | |:| |    ||||:|:|:||:|| |||||:|:| ||| |||:| ||:|| |:| |
Db    763 gaatggtagatctacacccaacgcttgcactgtgcacctgtgaaagccacagaaatcaat 822

Qy   1115 gycarccngtnaargcncngngnatggarw-sngtncntayaarcngtngtngcngar 1173
          |:| || ||:|||||:|:| | |||| |  |:|| ||||||:|:|:|| || |||||
Db    823 gccagaccgtgaaagcacctggtgctgtacctgcaaagccacaggggtggag 882

Qy   1174 ytnacnaaracngtngnathtayytnytncaygtgycaygayytngaygtnngncayggn 1233
          ||||| || | || | ||:||:||:||:|| |||:|||:|:|:||:||:||:|:| |||:|
Db    883 ctgccaaggccatgggagcccacgtcttgcatcagcatgacctggatgttagacatgga 942

Qy   1234 gtnaargmngaycayttygggnccnytnmngnttygaytgyccnacngngntttgmgnactay 1293
          |:|||| |:|| ||:||:|||:| ||:|:|:||:|| ||:|| ||:|:|:||:|:| |||:|
Db    943 gtcaaaggagatcaattggggccttaaaaattgactgcctactgatgttcagacttgc 1002

Qy   1294 atggggnccngt-nccnytntgytyggncarttyttyccnttgygnacngngntntyac 1352
          |||||:|| ||  ||:|:||:|:|:||| ||:||:||:||:|:| ||:|:|:|:|:||
Db   1003 gtgggccctgtaaccactttgtttggccaatatcccagttggaatggctgtattac 1062

Qy   1353 ncartgyttntayytncaytgyat 1376
          |:||:||:||:||:||:||:||:|
Db   1063 tcaatacctgtacctccactgcat 1086
```

```
RESULT 14
AAD05178
ID   AAD05178 standard; cDNA; 1278 BP.
XX
AC   AAD05178;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:68.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            224..268
FT                  /*tag= a
FT                  /product= "Human secreted protein"
FT                  /note= "CDS does not include start codon"
FT                  /partial
XX
PN   WO200134769-A2.
XX
PD   17-MAY-2001.
XX
```

PF 01-NOV-2000; 2000WO-US30040.
XX
PR 05-NOV-1999; 99US-0163580.
PR 30-JUN-2000; 2000US-0215130.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;
XX
PN WPI; 2001-308781/32.
DR P-PSDB; AAE01289.
XX
PT New isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 442; 519pp; English.
XX
CC AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC protein genes; and AAE01232-AAE01311 represent the proteins they encode.
CC AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 24 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify food storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1278 BP; 344 A; 264 C; 358 G; 299 T; 13 other;

Query Match 29.7%; Score 408.8; DB 22; Length 1278;
Best Local Similarity 45.8%; Pred. No. 4.8e-93;
Matches 458; Conservative 205; Mismatches 298; Indels 40; Gaps 6;

Qy 391 gargarathwsnaacarcarwsnathcargargtnacnctgggtnytnaargcnttyygtn 450
        |||||||::: | ||| ||::::|||:||:| | ||||||| ||::||| |||||||||
Db 280 gaagaaattcctaagcaagcaaagcattcaagaggtgattgggttgctgttaaaagcattt 339

Qy 451 wsnttyathmngargcngargcarcayaarwsnwsngaraayytncaycccngayaaygtnath 510
        |:|| |||::|||||||||||| |: |||||||||||||||||||:| ||||||||||:||
Db 340 catttaaagggaaac--agcataaaaatttggaaaatttgcaagccagttgatgcagca 397

Qy 511 aaraaraaaayccnttywsngarngnaarttyaaryytngcncngaarathtgyathtgy 570
        :| ||| |:| ||| ||:| |||||| | ||::|::||||:|||| | |||:::|:
Db 398 gaagagagaaaactgttttttggaggaagaaattcaagctggctgcagaaatttgcataagt 457

Qy 571 aaygargarytnaaaytcaaccncargayaayathwsntgacntgycar 630
        ::|: :||||||| |||||||:||||:|||||||||||||:|||| |:||
Db 458 tacaggaggctgaattgttaatctccaagacaatggggaaaatgtctccaaggccatgtcac 517

Qy 631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgncncmngmgnaar---tggtty 687
        :|:|||| || |||:||:||:| | ||:||||||:|||||:|:||:||||| | ||||
Db 518 aggtcttcattggcagccccctccatcacagacccctgaagactaggaggaaaaatggtt 577

Qy 688 tgyggnacnggncnggnwsnytntcgytgtygytgtncarccnmngnayyytngtncctgytgn 747
        ||:| || ||||:||||||||||||:|||:||||| |||||:||| |||:||||||:||
Db 578 tgtgggacaggagccagtcccatgctgtgtcagcctaggggactttggctgccctgcatc 637

Qy 748 ccngtnaaywsngcngtngc---------------------nwsngarg 775
        :||:| || ||:|||||||| |:||||
Db 638 ccagctgcttccagccaattgctaaaaggccgaggtacagctctgtcccatggttcagagg 697

Qy 776 gngcnwsnccnaarccntggcarytncccnwsngggtngarccncgttnggngcnaararw 835
        ||||:|||||| || |||||||||:|||||:||||:|| |:|||| |||| ||:| |||:
Db 698 gtgtcaagcgccccaaacccttgggcagctcccagtgggttggcaggcctgcaggtggaagt 757

Qy 836 snmngnathgargtngaarccnccncnathmgnttycaraaaraathtaygnaayccntgga 895
        :||||||| ||:||:|||:|| ||| || |||||:|:|| |||:||:|||:||||||||||
Db 758 caagaattcaggttgggaacctcatctagatttcaagaagatgtatggaatcacctgga 817

Qy 896 tgccnmgncarcaarttcygtcngtnggngtnggngnwsnwsntggmgnacnwsngcnmgngmgngtng 955
        ||| :||| || || ||:|:||:||:||||||||||:||||||||:|||| |||:|:|:||||
Db 818 tgctcaggcacaaaagtttgctccaggggcagagcctcatggagaacctctgctaggggcag 877

Qy 956 tncaraargtgnaaygtnggggarccncncaymgngtncccnwsngggngcnccnwsnw 1015
        |||:| ||||| ||:|||| |||||||:||| |||:|||||||||||:||:||||:|||
Db 878 tg-tgaaggaaatgtggggttcggagccccyacacagaaatcctactgggggccaccacta 936

Qy 1016 snmngngcngtnmngmngnwsnccncncnwsnwsnmgnytncaraargngmgnmgnwsnacngayw 1075
        :::||||:||||:|| ||||::|||:| ||:| | |||| |||||:||:|| |::||:|| ||
Db 937 gtggagctgtgcgagaaggccacctgtcctcagaccccagaatggtrgatccacccgaca 996

Qy 1076 snytncarcaygtncngaaraarwsnacncgayacnaartgycarccngtnaargcng-cn 1134
        :::|||||:| | |:||| ||:||||||:|||||||||||::||||||||||||||| |:
Db 997 gcttgtaccgwgcacctggaaaagccacagaccctaaagcaccgtgaaagcagcca 1056

Qy 1135 ggnatggarwsngtnccnntayaaracngtngtncgngarytnacnaaracngtngngnath 1194
        || ||||| |||||:|| |:|| |:| ||:|||:|||||:|||:||||||:|||:||||||
Db 1057 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaaggggaacc 1116

Qy 1195 tayytnytncaytgycaygayytngaygytngcncaygggntnaaringngaycaytygg n 1254
        |||:||||||||||||:|||||:||| |:| |:|||||:|||||:|||| ||:|||||||:
Db 1117 tacctcttgtcatcattgngacctgnatgtgaagacatggacatcaatttgga 1176

Qy 1255 gcnytnmgnttygaytgycnacnggnttymgnacntayatsggncncngtnaccnytntgy 1314
        || ||| |||||||:||||| ||||||||:| ||:|||| |||:|||||:|||:||:||||:
Db 1177 acgntataatttgactgctcgttggacttgcatggggcttgnaacgcttt 1236

Qy 1315 ttyg-gncarttyccnttygnacngcngcngttyacnc 1354
        ||:| ||||||||||:||||||||:||||||||:|||
Db 1237 gttcgtgcagttctcccattggaatgcgtgtgtattacc 1277

RESULT 15
AAS87262/c
ID AAS87262 standard; cDNA; 1580 BP.
XX
AC AAS87262;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23066.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

```
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23075.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PN   Claim 1; SEQ ID No 23066; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94554 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1580 BP; 357 A; 412 C; 328 G; 483 T; 0 other;
```

```
Query Match     29.1%;  Score 400.4;  DB 23;  Length 1580;
Best Local Similarity   42.4%;  Pred. No. 7.6e-91;
Matches  466;  Conservative  215;  Mismatches 383;  Indels  34;  Gaps  4;
```

```
Qy    307 aaytaytawsnwsnttygaygtngaryatncaycaytgyaargtnaayathtggytnatg 366
          ||:|| |:  |: |||||  ||  | |||: |||::|  || || ::||||| |||| |||
Db   1107 AATTCCGATAACGAACAAGACTGTGGGCATGCTAACTACTTACCTGAACCCCGAGCAGTCA 1048

Qy    367 mgnaaacaraathytngcnaayaargargarathwsnaarcarcarwsnathcaragargtn 426
          ||  ||| ||| |||| ||| || || || |||||:||  | || ||:|| | || || ||
Db   1047 GCATCCCCAACTTCTTAGAGGGACACAGTGGCATTCACCCACTCGAGATTGACCAATAAC 988

Qy    427 acntgggtnytnaargcnttywsnttyathmgnargcngarcyaarwsnwsngar 486
          || ||||| |||  || :||||:| ||| :|| || | || | |||||| || |||
Db    987 AGATGTGGTTTGGAATTGGAACTTATATTTATTTAAAAGGAAGCAGAGCCATAAAAATTGGA 928

Qy    487 aaytncaycccngayaygtnathaaraaararaayccnttywsngargngnaarttyaar 546
          || || |:|  || ||| :|| | ||  | ||  |||| |||:||:||:|| |||||:|||
Db    927 AATCTGCAGCATGACAGTGCAATAGACAAAAAAGAAAAACCCCATTTTCTGGGGAGAAATTCAAG 868

Qy    547 ytngcngcngarathtgyathtgyayayagyarygaryatatmgnaaygtnaaycncaragayaayggn 606
          :||| | | || || || ||| |:||  | |  |  || |:|| || |:|| | || || ||
Db    867 CTGTTGCAGAAATTTGCATAAATAACAAGGAGAACGCCAAATGTTTAATCACCAAGACAACGG 808

Qy    607 garaayathwsnttggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntgg 666
          ||| || |||| :| | ||: | ||| | || || ||  :|| || || || || |||
Db    807 AAAAATGTTTTCAGGGCACGTCAGAGACCCTTATGCCAGGACCCTCCCATCGACGAGCCCAG 748

Qy    667 mgnccnacngmgnaa---rtggttytgyggnacnggnccnggnwsnsnytntggytggytncar 723
          :|||| || || ||     ||| ||| :||||||| ||||:: || ||| ||  | |||||||
Db    747 AGGCCTACGAGGAGAAAAATATGGTTTTGTGGGCTCGGCCCCAGGACCTTGCTGCTTTGTGCAG 688

Qy    724 ccnmgngayytngtnccngtnaaywsng---------------------------------- 760
          ||  ||| || |||  | || ||||:||
```

```
Db    687 TCTCAGGACTTGGTGCCCTGGCCATCCCATCTGTGGCTAAAAGGGGCCAATATACAGCTGAA 628

Qy    761 -cngtngcnwsngargggngcnwsnccnaarccntggcaryytnccnwsnggngtngarccn 819
          ||  :||||  ||  | | || ||| || ||::| || | |:|| || || :|| |||:|
Db    627 ACCAATTGCTTCAGAGGGGTGTAAGCCCCAAGCATTGGGAGGCCTTACAAGTGTGGTGTTGGACCT 568

Qy    820 gtnggngncnaaraarwsnmgnathgargtntgggarcncnccnathmgnttycaraarath 879
          :: || | ||| | || |||| || ||| || || ||  ||| || | ||  |||||| |
Db    567 GTGGGTACACAGAAGTCAAGAAACTGAGGTTTGGGAAACCTCTGCCTAGATTTCAGAGGATG 508

Qy    880 tayggnaayccntggatgccnmgncaaraarttygcngtnggngtnggnwsnwsnggmgn 939
          ||  || || || || || || ::| ||| || || | :|| :|| || || || ||
Db    507 TATGGAAACACCTGGATATCCAGGCAGAGAGGTGTCGCAGGGGTGCGGCCCTCATGGAGA 448

Qy    940 acnwsngcnmgngtngtncaaraargnaaygtnggntgggarcncncnccaymgngtncn 999
          || ||||| || :|| || || || | || ||| | || | | |: |||| |||| || |
Db    447 ACCCTGCTGCTAGGGCAATGCAAGAGGAAAATGTGGGGTTGGAGCCTTCCACACAGAGTCCCC 388

Qy   1000 wsnggngcncncnwsnmgncngtnmgnamgnwsnccncncnwsnwsnmgnytncaraar 1059
          ||  || | |:|  ||  || || || || || || || | |  |:|| |||| ||||:
Db    387 ACTGG-----GACTAGTGGAGCGTGTGAGGAGGGGCCACCATCCTCCAGACCCCAGAAT 334

Qy   1060 ggngmgnwsnacngaywsnytncarcaygtnccngararaarwsnacngayacncaartgycar 1119
          || | || ||: || || ||| |||| || || | || || |||| | ||| || ||| ||
Db    333 GATAGAACCAACTGACAGCTTGGATCGTGTTGTTACCTGGAAAAAAGCCACAAAACACTCAATGCCAT 274

Qy   1120 ccngtnaargcngcnggngnatggacwsngtnccnctayaaracngtngtthcngarytnacn 1179
          || || ||: || | ||: | ||| |||: ||| ||: ||:|||| ||| ||| |||| |||
Db    273 CCTGTGAAAGCAGCCAGGAGGACCTGTACCCTCCAAACCCACAGGGGCGAGAGCTTCC 214

Qy   1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
          || | ||| || || ||: ||| |||| | ||:|| ||:|| | | ||: ||: || ||
Db    213 AAGGCTGTAGAAAACCCACTCTTACATCAGCGTGTTCCTGGATGTGAGACCTGGGAGTCAAA 154

Qy   1240 mgngaycaytyggngcnytnmgnttygaytgyccnacnggngttymgnacntayatgggn 1299
          :||| | || ||| |||   | |||| ||:|| ||| || |||  | | ||  || |||
Db    153 GGAGAATCATTTGGAGTTTTAAGATTTGCCGGTGCCCCACTGGATTTGGAGTTGCATGGGG 94

Qy   1300 ccngt-ncnaytnttgytyggncarttytyccnttyggnacncgngtntttyacncartg 1358
          || :|  :| || ||:|| ||:|| ||:||  ||:|| :|| || ||:| | ||: ||
Db     93 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATCTGGAATGAGTGTGTTTATCCAATG 34

Qy   1359 yytntayytncaytgyat 1376
          :: ||| ||||| :|:||
Db     33 CCCGTACCCCCATTATAT 16
```

```
Search completed: July 8, 2002, 16:20:51
Job time: 12424 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 12:57:56 ; Search time 104.62 Seconds
                                              (without alignments)
                                              3233.008 Million cell updates/sec

Title: US-09-997-610-3
Perfect score: 1377
Sequence: 1 athgtngtnathccngtnyt..........gyytntayytncaytgyatg 1377

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0 .

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   Issued_Patents_NA.*
                  1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                  2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                  3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                  4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                  5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                  6: /cgn2_6/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1	c	299.2	21.7	49136	4	US-09-422-869-1	Sequence 1, Appli
2	c	177.6	12.9	14855	2	US-08-687-080-59	Sequence 59, Appl
3	c	164.6	12.0	620	4	US-08-991-789A-29	Sequence 29, Appl
4	c	164.6	12.0	620	4	US-09-062-451-29	Sequence 29, Appl
5	c	140.6	10.2	152331	3	US-09-128-155-16	Sequence 16, Appl
6	c	135.8	9.9	14855	2	US-08-687-080-59	Sequence 59, Appl
7	c	67	4.9	7218	1	US-08-232-463-14	Sequence 14, Appl
8	c	59	4.3	2067	4	US-09-106-194-11	Sequence 11, Appl
9	c	51.8	3.8	1313	2	US-09-140-804-9	Sequence 9, Appli
10	c	51.8	3.8	4517	4	US-09-140-804-9	Sequence 9, Appli
11	c	50.6	3.7	1155	3	US-09-053-866-3	Sequence 3, Appli
12	c	48.8	3.5	2082	4	US-09-440-325A-2	Sequence 2, Appli
13	c	48.8	3.5	2088	4	US-09-351-414-3	Sequence 3, Appli
14	c	48.6	3.5	289	4	US-09-007-005-17	Sequence 17, Appl
15	c	48.6	3.5	289	4	US-09-244-796-17	Sequence 17, App
16	c	48.6	3.5	1107	3	US-09-188-930-217	Sequence 217, App
17	c	48.6	3.5	1276	2	US-08-463-911-1	Sequence 1, Appli
18	c	48.2	3.5	2277	1	US-08-676-967-2	Sequence 2, Appli
19	c	48.2	3.5	2277	1	US-08-676-974-2	Sequence 2, Appli
20	c	48.2	3.5	2277	2	US-09-098-487-2	Sequence 2, Appli
21	c	48	3.5	2277	2	US-08-676-974-2	Sequence 2, Appli
22	c	48	3.5	2277	1	US-09-098-487-2	Sequence 2, Appli
23	c	48	3.5	2277	1	US-09-098-974-2	Sequence 2, Appli
24	c	44.4	3.2	1212	4	US-09-092-770-18	Sequence 18, Appl
25	c	44.4	3.2	1212	4	US-09-222-851-18	Sequence 18, Appl
26	c	43.4	3.2	5181	1	US-08-257-073-10	Sequence 10, Appl
27	c	43	3.1	1215	2	US-09-092-770-8	Sequence 8, Appli
28	c	43	3.1	1215	4	US-09-222-851-8	Sequence 8, Appli
29	c	42.4	3.1	1015	3	US-09-188-930-30	Sequence 30, Appl
30	c	42	3.1	2067	4	US-09-106-194-11	Sequence 11, Appl
31	c	41.8	3.0	1001	3	US-09-188-930-218	Sequence 218, App
32	c	41.8	3.0	2265	3	US-09-369-618-3	Sequence 3, Appli
33	c	41.8	3.0	2265	3	US-08-934-959-7	Sequence 7, Appli
34	c	41.6	3.0	510	2	US-08-934-959-7	Sequence 7, Appli
35	c	41	3.0	597	3	US-09-528-760A-3	Sequence 3, Appli
36	c	40.8	3.0	723	3	US-08-911-423-5	Sequence 5, Appli
37	c	40.8	3.0	1659	1	US-08-231-729B-2	Sequence 2, Appli
38	c	40.6	2.9	2949	4	US-09-412-554A-3	Sequence 3, Appli
39	c	40.2	2.9	500	3	US-09-141-000-2	Sequence 2, Appli
40	c	40.2	2.9	1559	1	US-08-231-729B-1	Sequence 1, Appli
41	c	40	2.9	2088	4	US-09-351-414-3	Sequence 3, Appli
42	c	39.8	2.9	2265	3	US-09-369-618-3	Sequence 3, Appli
43	c	39.8	2.9	2265	1	US-09-369-617-3	Sequence 3, Appli
44	c	39.6	2.9	4379	1	US-08-592-214A-17	Sequence 17, Appl
45	c	39.6	2.9	4379	3	US-09-149-976-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-422-869-1
; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

Query Match   21.7%; Score 299.2; DB 4; Length 49136;
Best Local Similarity   41.1%; Pred. No. 1.1e-72;
Matches 388; Conservative 186; Mismatches 307; Indels 62; Gaps 5;

Qy   444 rgcnttywsnttyathmgngargcngarcayaarwsnwsngaraaaytncaycccngayaa 503
            :||::|:|| ||:: |:| ||::|| :|||:: |:||::: :|: | ||:||::|:: |
Db   47384 ggaactcacatcatttaaaagggaagcaagcataaaagttagaaaaattgcatcctgacta 47443

Qy   504 ygtnathaaaraaaraaaaaycccnttywsngarcggnaarttyaarytngcncngarathtg 563
            | ||| |:: :|:: | |:|| |:| ||:|| ||| | ||:|| | :| | ::|||| :||
Db   47444 tgggatagaaaaagaaaaaactcattttctgagagaaatttcaagccacgctgcagaaattg 47503

Qy   564 yathtgyaaygargaryytnaaygtnaayccncargayaayggngaraaayathwsnttgac 623
            :||:|||:|| :|| | |:||:||::| |:|:| |:||| | ||:| ||:|||: ||:
Db   47504 cataagtaactaggagcccacatgttaatagacatagacaatgaggaaatgtctccagggc 47563

Qy   624 ntgycarngnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg--naa 680
            :||:||:|| ||:|||:||:||:| | |:|| | ||:|:| |: || |||:|:|: naa
Db   47564 atgtcagaggtcttcacagcaacccacccatcacaggcctcggagcgcttagaggagaaa 47623

Qy   681 rtggttytgtgygnacngnccnggnwsnytntgytgygtncarccnmgngayytngtncc 740
            :||:||:|:|::|| ||:|::| ||||:||:||:||:||:|:|||::| |||:||:||:||

```
Db 47624 atggttttgtcgttgggcccaaggccttgcttgcgtttgtgcaagtccaaggaccttggtgcc 47683
          :||||||||:||:| |: || | :||| | ||| | || | || |:| : | ||| || ||
Qy  741 ntgygtnccngtnaaywsngngcngtngcnwsnccnaarccntggccaryt         800

Db 47684 ccacatcc------cagcaggtgctaaaagggccaatgtacagcttagacctt 47731
          ::|| |         | |||| | | || || | | |||| ||  |
Qy  801 nccnwsnggngntngatccngtnggngngcnaaraaarwsnmgnathgargtnt ggaccncc 860

Db 47732 tgcttcagaggtgcaagc------cc 47751
          |:||| | ||| |:|                        |
Qy  861 nathmgnttycaaraarathtayggnaaycntggatgccnmgncaraaarttygcngtngg 920

Db 47752 caagccttggtgcttacatgtggttgggcctgcagatacacagaaagtttgctgcact 47811
          | || | || |:|| | | :|||:|| :|||| | ||| | || |:||| ||:|
Qy  921 ngtnggnwsnwsnsntggmgnacnwsnmgnmgntngtncaraaarggnaaygtnggntngga 980

Db 47812 ggtggaaaccctcatgtagaaacctctgctaggggcagtgtagaagtgatatgtgggttgga 47871
          ||:|| | || || | ||| | | |: || || | | ||| | ||: | | ||:|||
Qy  981 rccnccncaymgngtnccnwsnggngncnc------cnwsnwsnmgncngtnmgnmgnws 1034

Db 47872 gccccccacacaatccccactgggncaactcccactgctactgtgctactgtgagaagaag 47931
          || | | || | :||| || | || ::| |:| || || |: |||
Qy 1035 ncccncnwsnwsnmqnnytncaraarggnmmwsnacngaywsnytncaraaytgnccnga 1094

Db 47932 gccacatccctccagacnacccagaatggtagatccattccactggcttgaacctgcacctgg 47991
          || ||:|:||::||| | : || || | ||| | | |||:|| | ||| ||
Qy 1095 raarwsnacngayacnacaartgycarccngtnaargcngcnggnatggarwsngtnccnta 1154

Db 47992 aaaagcccacagacactcaacacacagccctgaaggcagctgaaggagcctgaagctgaacctg 48051
          ||: | || | :|||| | || |||||| ||| :|| |:|| || |
Qy 1155 yaaracngtngtncngaarytnacnacracngtnggnathtayytnaytgycaygga 1214

Db 48052 caaaacaacagaggcagagagcgccaaggtcatggagagcccaccttgcatgagcctga 48111
          | || | ||| | | || | || |||| | :|: | | :||| |::|||||
Qy 1215 yytngaygtnmgncaygggntnaarmgngaycaytnggyggnccnytnmgnttygaytgycc 1274

Db 48112 cttgaatgtgagacaggagtcaaaggagatcaatttaagtttgtgactgccc 48171
          |:||| | ||| | || |||| |||| |::|| || | |||| || |||
Qy 1275 nacnggnttymgnacntayatggngnccngt-ncnnytntgyttygggncaarttytyccnt 1333

Db 48172 acctggattcggacttcgcatggggcctgttggccctcatttggccaatttatccat 48231
          || |:|| | || || | |||| |:| ||| || ||| || ||
Qy 1334 tyggnacngcngtnttyacncaartgyytntatayytncaytgyat 1376

Db 48232 ttggaatgggtatattaccaatgcctgtaccccattctat 48274
```

```
Query Match       12.9%;  Score 177.6;  DB 2;  Length 14855;
Best Local Similarity  44.4%;  Pred. No. 7.7e-39;
Matches 221;  Conservative 90;  Mismatches 172;  Indels 15;  Gaps 4;

Qy  890 cntggatgccnmgncaraaarttygcngtngtnggngnwsnwsntggmgnacnwsngcnm 949
         |||||| ||| || |||| || || || ::| :|| |: ::|| | || |
Db 14854 CCTGGAGTCCCAGGCAGATAGTTTGCTCGCAGGGGTGGGCCCTTATGGAGAAACCTCTGCTA 14795

Qy  950 gngtngttncaraarggnaaygtnggntnggarccncncacymgngtnccnwsnggngncnc 1009
         |: |:|| |||| ||| || || ||:|: || | | || | : || | || |
Db 14794 GGGCAAATATGGAAGGGAAATGTGGGGTTGACTGAAAACCCAC--AGAGTTCCTATGGAGGGAC 14737

Qy 1010 cnwsnwsnmgncngtnmgnmgnwsncnccnwsnwsnmgnyt--------ncaraar 1059
         |: || | | |:|| | || || | | |:| :|:       | |||
Db 14736 TGCCTAGTGGAGCTGTGAGAAGACAGCCACTGTCCTCCAGACTGGTAGATCCCCAGAAT 14677

Qy 1060 ggnmgnwsnacngaywsnytncaraagtnccngaraarwsnacngayacnacaartgycar 1119
         || :| || || || : || || ||| | :|||||| ||:|| || | |
Db 14676 AATAGATCCACTGACACAGCTTGCACTGTGCACCTGGAAAAACTGCAGGCGCACTCAACACCAG 14617

Qy 1120 ccngtnaargcngcnggnatggarwsngtnccncatayaaracngtngtncngaarytnacn 1179
         || |:|||:|:|| || ||||| | |||||| | |: ||| | || | |: ||
Db 14616 CCTGTGAAAAACAGCCAGGAAGGAGGCTATACCCTGCAAAGCC--AGAAGTGGAGCTGCCC 14559

Qy 1180 aaracngtnggnathtayytncaytgycaygaaytgnaygtnmgncayggngtnaar 1239
         || |||:|||| | ||:|| |||| |||| :| :|||| || |||| | |||
Db 14558 AAGGCCATGGAAGGCCCACCTCTTGCATCACGAGTGACCTGGATGTGAGACACATGGAGTCAAA 14499

Qy 1240 mgngaycaytyggngncytnmgntyggaytgycnacngnttymgnacntayatgggn 1299
         || |:|||:| || || ||| | ||| || | ||:|| || || || ||
Db 14498 GGAGATCAATTCTCGGAGCTTTAAGATAACACCTGCCCACTGAATTTCGGACTTGCACGGGG 14439

Qy 1300 ccngt-nccnytntgyttyggggncaarttyttyccnttyggnacngcngtnttyacncartg 1358
         || |  ||| |:|| || || |||||| :|||| ||| |||| ||| | ||||:| | ||
Db 14438 CCTGTAGCCCCTTGTTTTGGCCAATTTCTTCCATTTTGGAATGGCTTGTATTTGCCCAATG 14379

Qy 1359 yytntayytncaytgyat 1376
         :| |||:|| |||:|||:
Db 14378 CCTGTATCCCCATTGTAT 14361
```

```
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
    TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
    ADDRESSEE: Seed IP Law Group
    STREET: 701 Fifth Avenue, Suite 6300
    CITY: Seattle
    STATE: Washington
    COUNTRY: USA
    ZIP: 98104-7092
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/991,789A
    FILING DATE: 11-Dec-1997
    CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
    NAME: Potter, Jane E. R.
    REGISTRATION NUMBER: 33,332
    REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (206) 622-4900
    TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
    LENGTH: 620 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-08-991-789A-29

Query Match       12.0%;  Score 164.6;  DB 4;  Length 620;
Best Local Similarity  40.7%;  Pred. No. 1.6e-36;
Matches  250;  Conservative 105;  Mismatches 243;  Indels 17;  Gaps 6;

Qy  772 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar 831
        |:||:|| || |||||||| |||::| ||||||| || |||| || |||| |||
Db    3 GAGGGTACCAGCCCCAAGCCTTGACACACTTCCATAGGGTGTCAAGCCTGTGGGTGCACAG 62

Qy  832 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtaygg------- 884
        |||:|| ||| |||| |||||| ||| || |||||||| |:| ||||||
Db   63 AAGTCAAAAATTGAGTTTTGGGATCCCACCTAGATTTCAGAGGATATAAAAGAAACACC 122

Qy  885 naayccntggatgccnmgncaraarttygcngtngngtnggnwsnwsntggmgnacnws 944
        ||| || ||| ||| || |||||| ||||| |||| || :|| :|
Db  123 TAACACCTAGATATTCAGACAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC 182

Qy  945 ngcnmgntngtncaraarggnaaygtnggngtnggngtnccnwsngg 1004
        || ||| || ::|| :|
Db  183 TACTAGGAAAAGTACAGAAGAGAAAATGTGGGTTTGGAGCCCCAAAACAGAATCCCTCTAG 242

Qy 1005 ngcnccnwsnsmgngncgngtnmgnmgnwsncncnwswsnmgnytncaraarggnng 1064
        |||| ||::| | |||| |::| ||
Db  243 AACACTGCCTAATGAAAACTGTGAGAAGAATGGCCACTGTCATCCAGACACCAGAATGATAG 302

Qy 1065 nwsnacngaywsnytncarcayg-tnccngaraarwsnacngayacncartgycarcc-- 1121
        |||| || ::||| || | | || || ||||:| || || || |
Db  303 ACCCACCACAAAAACTTATGCCTATAAAACTACAGACACTCAATGCCAGCCCC 362

Qy 1122 -ngtnaargcngcngnatggarwsngtnccntayaaracngtngtngcngarytnacna 1180
        || ||||:|| || || |:||| || || | |||| ||
Db  363 ATGAAAAAAAAACTGAGAAGAGACTGTNCCCTACAATGCCACGGAGCAGAACTGCCC 422

Qy 1181 aracngtnggnathtayytnytncaytgycaygayytngayg-tnmgncaygggtnaar 1239
        || | | |::||| | |:| || :|| |
Db  423 AGGCCATGGAAGCCACAGCTCTCTATATCAAATGTGACCTGGATGTTGAGACATGGAATCCNA 482

Qy 1240 mgnggaycayttyggngcnytnmgntt----ygaytgyccnacngngntctymgmgnacntayat 1295
```

```
Db  483 NGAAATCNTTTTAANACTTCCACGGTTNAATGACTGCCCTATTANATTCNGAACTTANAT 542
        |:|| |||| ||::| | |||||| || |||| ||
Qy 1296 g-ggnccngtnccnoytntgyttyggncarttytyccnttgyggnacngngtnttyacnc 1354
        |||:|| ||| |||| |||||| ||:|| || ||:|| :|
Db  543 CCNGGCCTGTGACCTCTTTCGCCATTCCCCCCTTTTTGGAATGGCTNTTTTTTTTCC 602
Qy 1355 artgyytntayytnc 1369
        ||:||:| ::|
Db  603 CATGCCTGTNCCCTC 617
```

```
RESULT  4
US-09-062-451-29
Sequence 29, Application US/09062451
Patent No. 6344550
GENERAL INFORMATION:
    APPLICANT: Frudakis, Tony N.
    APPLICANT: Smith, John M.
    APPLICANT: Reed, Steven G.
    TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
    TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF BREAST CANCER
    NUMBER OF SEQUENCES: 297
CORRESPONDENCE ADDRESS:
    ADDRESSEE: SEED and BERRY LLP
    STREET: 6300 Columbia Center, 701 Fifth Avenue
    CITY: Seattle
    STATE: Washington
    COUNTRY: USA
    ZIP: 98104-7092
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/062,451
    FILING DATE: 04-APR-1997
    CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
    NAME: Maki, David J.
    REGISTRATION NUMBER: 31,392
    REFERENCE/DOCKET NUMBER: 210121.419C2
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (206) 622-4900
    TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
    LENGTH: 620 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
US-09-062-451-29

Query Match       12.0%;  Score 164.6;  DB 4;  Length 620;
Best Local Similarity  40.7%;  Pred. No. 1.6e-36;
Matches  250;  Conservative 105;  Mismatches 243;  Indels 17;  Gaps 6;

Qy  772 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar 831
        |:||:|| || |||||||| |||::| ||||||| || |||| || |||| |||
Db    3 GAGGGTACCAGCCCCAAGCCTTGACACACTTCCATAGGGTGTCAAGCCTGTGGGTGCACAG 62

Qy  832 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtaygg------- 884
        |||:|| ||| |||| |||||| ||| || |||||||| |:| ||||||
Db   63 AAGTCAAAAATTGAGTTTTGGGATCCCACCTAGATTTCAGAGGATATAAAAGAAACACC 122

Qy  885 naayccntggatgccnmgncaraarttygcngtngngtnggnwsnwsntggmgnacnws 944
        ||| || ||| ||| || |||||| ||||| |||| || :|| :|
Db  123 TAACACCTAGATATTCAGACAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC 182

Qy  945 ngcnmgntngtncaraarggnaaygtnggngtnggngtnccnwsngg 1004
        || ||| || ::|| :|
```

RESULT 5
US-09-128-155-16
Sequence 16, Application US/09128155
Patent No. 6177654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 0404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 10.2%; Score 140.6; DB 3; Length 152331;
Best Local Similarity 32.8%; Pred. No. 2.1e-27;
Matches 324; Conservative 168; Mismatches 426; Indels 71; Gaps 6;

```
Db   183 TACTAGGAAAGTACAGAGAGAAAATGTGGGTTGGAGCCCCAAACAGAATCCCCTCTAG 242
Qy  1005 ngcnccwsnwsnmgngncgntmgnamgnwsnccncnwsnmgngnytncaraarggnmg 1064

Db   243 AACACTGCCTAATGTAAACTGTGAGAAGATGGCCACTGTCATCCAGACACCAGAATGATAG 302
Qy  1065 nwsnacngaywsnytncarcayg-tncngaraarwsnacngaycacnncarttgycarcc-- 1121

Db   303 ACCCACCAAAAAACTTATACCATATAAAAACTACACAGACACTCAATGCCAGCCCC 362
Qy  1122 -ngtnaargacngcnggngatggaarsngtnccntayaaracngtngtngcngarytnacna 1180

Db   363 ATGAAAAAAAAAACTGAGAAGAAGACTGTNCCCTACAATGCCACCGGAGCGAGAACTGCCCC 422
Qy  1181 aracngtnggnathtayytnytncaytgycaygayytnga yg-tnmgncaygggngtnaar 1239

Db   423 AGGCCATGAGAAGCACAGCTCTTATTCATCAATGCCTGGATGTTGAGACATGGAATCCNA 482
Qy  1240 mgngacaytgggngcnytnmgntt---ygaytgyccnacnggnttymgnacntayat 1295

Db   483 NGAAATCNTTTTAANACTTCCACGGTTNAATGACTGCGCCTATTANATTCNGAACTTANAT 542
Qy  1296 g-ggncngtnccnytntgyttyggncartty ytyccnttyggnacngcngtnttyacnc 1354

Db   543 CCNGGCCTGTGACCTTCTTTGCTTTGGCCATTCCCCTTTTTGGAATGGCTNTTTTTTTTCC 602
Qy  1355 artgyytntayytnc 1369

Db   603 CATCCCTGTNCCCTC 617
```

RESULT 6
US-08-687-080-59
Sequence 59, Application US/08687080
Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

```
Db 71345 gcaatcattgctag---agagattagcatgactaaaaggagccaagtgctaattca 71400
Qy   597 rgayaayggngaraayathwsntgyacyntgycarmgnwsnwsncarcarwsnathaarws 656

Db 71401 agacaatgttaaaagggccttgaggcattcagagatctatgaagcgccctccatc 71460
Qy   657 nytngcntggmg-------nccnmqmgnaartggttyggnacngngnccnggnwsny 709

Db 71461 acagtgcagaggttggtgcactaggcccaagagagtttatgggactnacngnacngggcac 71520
Qy   710 tntgytgygtncaarccnmgngayytngtnccntgytgycngtnaa--------- 755

Db 71521 actgctatgcacagcttgggacactgctgcccgnacttctggctcc 71580
Qy   756 ------------------ywsngcngtngcnwsngargngngcnwsncc 785

Db 71581 acccctgctcaaacgggccaagatagaagcttgaccactgctccagaggccaagcca 71640
Qy   786 naarccntggcarytnccnwsngggngtngarccnctnggngcnaaraarwsnmgnmath-- 843

Db 71641 taagcctgyggtgttccatgtgtttaagcctgcaggtgccaagaatgcaagattgag 71700
Qy   844 gargtntggarccnccnathmgnttycaraarathtaygngnaayccntggatgtgccnmgn 903

Db 71701 gggagcttggggcaccttccacctaaattcagaggatgtgtcagaaacccntaggttcccagg 71760
Qy   904 caraarttygcngtnggngtnggnwsnsntggmgnacnwsngcnmgntngtncaaraar 963

Db 71761 cagaagcatgatacagggngggcagagccctggcaagctaggggcaatgccaaag 71820
Qy   964 ggnaaygtnggntgggarccncnccnacaymgntncnwsngngcncnwsnwsnmgngcn 1023

Db 71821 gaaaatgtgggtgttgggagtcctcacacatgtcccacctgggggcaactacctggtgatact 71880
Qy  1024 gtnmgngmgnwsnccncncnwsnwsnmgngnytncararggnmgnwsnwsnacngaywsnytncar 1083

Db 71881 gtgggaatgggctgctgccctcccagaatggtcacatggcagctggcac 71940
Qy  1084 caygtnccngararaarwsnacngayacncartgycarccngtnaargcncnggnatggar 1143

Db 71941 cctgagcctggaaagctgcagcactcaacctccaacccatgagatcaagccacatgggct 72000
Qy  1144 wsngtncentayaaracngtngtngcngaaytnacnaaracngtnggnathtayytnytn 1203

Db 72001 ac--tccaggagccccacagaggcagggctgctaagcctacccctg 72058
Qy  1204 caytgycaygaytngaygtnmgncayggngtnaarmgnqaycayttyggngcnytnmgn 1263

Db 72059 aaccgctgcaggacatggaatcaa------agattatgttgcagcttaagg 72106
Qy  1264 ttygaytgyccnacnggnttymgnacntayatgggncngtnccnqtntgyttyggncar 1323

Db 72107 cttaatgtttccctgtcaattcaggctgtgtgggacctgttgcttttttttttt 72166
Qy  1324 ttyttyccnttyggnacngcngtgtnttyac 1352

Db 72167 ttttttttttttggtcacaggtgtttgaac 72195
```

```
;  ZIP: 94306
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/687,080
;   FILING DATE: 17-JUL-1996
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/592,126
;   FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Sholtz, Charles K.
;   REGISTRATION NUMBER: 38,615
;   REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 324-0880
;   TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 59:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 14855 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     INDIVIDUAL ISOLATE: 5' END OF INTRON 2 OF RAD50 GENOMIC
; US-08-687-080-59
;     INDIVIDUAL ISOLATE: SEQUENCE

Query Match    9.9%;  Score 135.8;  DB 2;  Length 14855;
Best Local Similarity  39.5%;  Pred. No. 4.9e-27;
Matches 191;  Conservative 101;  Mismatches 160;  Indels 31;  Gaps 3;

Qy   444 rgcnttywsnttyathmngnargngcngarcayaarwsnwsngaraayytncayccngayaa   503
           ||:|| ||| :|:|| |||:|:|| || | |:|::|| |   |::|| || :| | ||
Db  5187 GGAACTTATATTTAAAAGGGAAGACAGAGCATAAAAAGTTTGAAAATTTGCACCCTGATCA  5246

Qy   504 ygtnathaaraaraaayccnttywsngargngnaarttyaaarytngcngnarathtg      563
          || ||: || :| || | ||| ||:|| ||:|:|| ::|||:|:| |:| | |:| |
Db  5247 TGTGGTAGAAAAAGAAAAACCCATTTTCTGGGGAGGAATTCAAGCCGGCTGCAGAAATTTG  5306

Qy   564 yathtgyaaygargarytnaayg-tnaayccncargayaayggngaraayathwsntgga   622
          :| | ||:||:||:| |:||::|  |:|  :|:|| ||:|| || ||:|| |:|:|:||
Db  5307 CATAAGTAAAGAGGAGCAGAATGTTTAATAGCCAAGACACTGGGGAAAATGGCTAGAAGG   5366

Qy   623 cntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaart   682
          |:|| || :|:|::|:| |||| |:||: |: ||:|:| ||:|:|  |:| :|::|| |
Db  5367 CATTTCAGAGATCTTCACACCAGCCCCTCCCATTACAGGACACTACTCCC----------  5416

Qy   683 ggttcytgyggnacnggncnggnwsnytntcgytgytgtncarccnmgnayytngtnccnt  742
          ||||: || :| || |:|:||:|:|| | |: |:|| ||:||| : |::|:||||:|| |
Db  5417 TGTGTTTCAGCCACTCCAGCTCCAGCCATGGCTAAAAGGGCCCCAGATATATC------   5469

Qy   743 gygtnccngtnaaywsngcngtngcnwsngargngncnwsnccnaarccntggcarytnc   802
          :| | | |:|| |:|:|   || |  |:|| ||:|:||:|| :| || :||||| ::|
Db  5470 -------TTAAGTGCTGCTCCAAGCCTGCAACCTGTAAGCCTTGGCAGCTTC         5516

Qy   803 cnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntggarccnacna   862
          ::||::|:|| || |:|||||:|: |||| || ::| |: |:|:| || | |:| ||
Db  5517 CACACGGTGTTAAGCCTCGGGGTTGCCAGAGAACAAGAGTTGAGGCTTGGGAGCCTCCAC  5576

Qy   863 thmgnttycaraaarathtayggnaayccntggarccncacna   922
          :::| ||:|| ||:|| ||||||:|:||| | |:|| || |
Db  5577 CTAGATTTCAGGAGGATGTATGGAAAGCTCTGGATCCAGGACTTG  5636

Qy   923 tng   925
          | |
Db  5577 CTAGATTTCAGGAGGATGTATGGAAAGCTCTGGATCCAGGACTTG  5636
```

```
Db  5637 TTG  5639

RESULT 7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
; CURRENT APPLICATION DATA:
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 1800 Diagonal Road, Suite 500
;   CITY: Alexandria
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22313-0299
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/232,463
; PRIOR APPLICATION DATA:
;   CLASSIFICATION: 435
;   APPLICATION NUMBER: US/07/935,313
;   FILING DATE: 26-AUG-1991
;   APPLICATION NUMBER: EP 91 114 300.6 .
; ATTORNEY/AGENT INFORMATION:
;   NAME: BENT, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)683-4109
;   TELEFAX: (703)836-9300
;   TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTzgpt-Pls
; US-08-232-463-14

Query Match    4.9%;  Score 67;  DB 1;  Length 7218;
Best Local Similarity  12.6%;  Pred. No. 6.4e-08;
Matches  64;  Conservative 203;  Mismatches 241;  Indels 0;  Gaps 0;

Qy   226 athathttyacnggngtnyttntayaaygcncarmngngayytnaargcnatggngngtn  285
          || :: ||:||||:|| | :||:||:|| |:|| :||:|||:||:|| ||:|: | ::|
Db  1536 ACGCTCTAGACGCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTA  1477

Qy   286 ttygcntgymgngtncnggnaaytaytayaywsnwsnttygaygtngarytncaycaytgy  345
          ||:|:|:|:|:|||:|:|||||:||:||:|:|:|:||||:|:||:||::|||| | :|||
Db  1476 CCTATCTATGCAAGTAGTAGTTAAAGAGATAGAAGAATTGGTACRRRRRRRRRRRRRR   1417

Qy   346 aargtnaayathtggytnaatgmgnaarcarathytngcnaayaaargargarathwsnaar  405
          || |:|| |:| || | |:|:|:|:|| || |: |:|:| ||:|| | |: || | |:||
Db  1416 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR   1357

Qy   406 carcarwsnathcargargtnacntgggtnytnaargcnttywsnttyathmngngar  465
          ||:||:|| |: ||:||:|: | |:|||:| |:||||| ||:|:| ||:|:|:|:||
Db  1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR   1297
```

```
Qy   466 gcngarcayaarwsnwsngaraayytncayccngayaayytnathaaraaraayccn 525
         ::|:|  |:: |:|| ::|| |:|:|:|| ::|::|::: |:| |::|::|::|
Db  1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237

Qy   526 ttywsngargggnaarttyaarytngcngcngarathtgyathtgyaaygargartnaay 585
         ::|:|  |:: |:|| ::|| |:|:|:|| ::|::|::: |:| |::|::|::|
Db  1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177

Qy   586 gtnaayccncacgayaaygggngargaayathwsntggacntgycaarmgnwsnwsncarcar 645
         ::|:|  |:: |:|| ::|| |:|:|:|| ::|::|::: |:| |::|::|::|
Db  1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117

Qy   646 wsnathaarwsnytngcntggmgnccnngaynaaRrtggttytgyggnacnggnccnggn 705
         ::|:|  |:: |:|| ::|| |:|:|:|| ::|::|::: |:| |::|::|::|
Db  1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCT 1057

Qy   706 wsnytntgytgygtncarccncmngaayy 733
         ::|:|  |:: |:||  |:|  |:|
Db  1056 CCCTCGACCTGCAGCCAAGCTCGGAATT 1029
```

RESULT 8
US-09-106-194-11
Sequence 11, Application US/09106194
Patent No. 6262234

GENERAL INFORMATION:
   APPLICANT: Holloway, James
   APPLICANT: Jelinek, Laura
   APPLICANT: Durnam, Diane
   APPLICANT: Blumberg, Hal
   TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR POLYPEPTIDE
   TITLE OF INVENTION: ZPPAR4
   NUMBER OF SEQUENCES: 11
   CORRESPONDENCE ADDRESS:
   ADDRESSEE: ZymoGenetics, Inc.
   STREET: 1201 Eastlake Avenue East
   CITY: Seattle
   STATE: WA
   COUNTRY: USA
   ZIP: 98102
   COMPUTER READABLE FORM:
   MEDIUM TYPE: Diskette
   COMPUTER: IBM Compatible
   OPERATING SYSTEM: DOS
   SOFTWARE: FastSEQ for Windows Version 2.0
   CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/106,194
   FILING DATE:
   PRIOR APPLICATION DATA:
   CLASSIFICATION:
   APPLICATION NUMBER:
   FILING DATE:
   ATTORNEY/AGENT INFORMATION:
   NAME: Leith, Debra K
   REGISTRATION NUMBER: 32,619
   REFERENCE/DOCKET NUMBER: 96-11
   TELECOMMUNICATION INFORMATION:
   TELEPHONE: 206-442-6674
   TELEFAX: 206-442-6678
   TELEX:
   INFORMATION FOR SEQ ID NO: 11:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 2067 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: single
   TOPOLOGY: linear
US-09-106-194-11

Query Match   4.3%; Score 59; DB 4; Length 2067;
Best Local Similarity 9.3%; Pred. No. 3.5e-06;
Matches 101; Conservative 297; Mismatches 676; Indels 12; Gaps 3;

```
Qy    71 cnmgnccncargargtnggnccnccngghgnccnggngtntncncartayacnggng 130
         :  |   |:|: |:|:|  |::| ::| |:|  |::|  |:|:  |
Db    13 YTNGARMGNCCNCCNGGNCARGTNWSNYTNGCNCCNTTYTNCGNGCNTNNCARCAR 72

Qy   131 arathwsngaratgacnaartgyccntgyccngayathgarmgnwsngcnttyacngtna 190
         |:|:|  :| ||   ::|:| |  |:|  |:|:|| ::| :|| |::|
Db    73 AARTTYACNTYTNGGATHMGNCARGCNCARATGWSNTGYGARAAYGCNCAYGCNTNCCN 132

Qy   191 arytnwsnggnaaaryttnccnytncacycayacnathatnttyacngngngtntaya 250
         |:| |   |:| ||  ::|:|  ::||  ||:  |:| |   |::|    |
Db   133 WSNACNCNCCNCCNTTYTNYTNCNNAARCCNCNCNCCNCCNGCNTNCCN 186

Qy   251 aaygcncaarmgnagaytcnaatggngtntttygcntgygmgntgtnccnggnaayt 310
         |:|  |  :|::||  |::|| |:| |::|  |||  |:|
Db   187 WSNCCNWSNWSNNYTNYTNACNAARCCNCCNCCNCACNCCNCGNGCNCNYTNCCN 246

Qy   311 aytaywsnwsnttygavrgtngarytncaycaytgyaargtnaayathtgyytatgmgna 370
         |:| |:| |   |:| |||  |::  |:: |: :|   |  |:|
Db   247 CCNCCNCNCARWSNNAARWSNCARAARAAYSNWSNGARCCNGARGCNGTNGCNTTYTTY 306

Qy   371 arcarathytngcnaayaargaratathwsnaarcarcarwsnathcargargtnaacnt 430
         |:| |   :|:|  :| ::  |:|:|  |::  ::|  :|:  |
Db   307 AARAARCARGCNCAYTGGMGNGARMGNAARMGNAAAGNNCARCCNAARCARAAYCCNGGN 366

Qy   431 gggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraayy 490
         |:|  |  |:|   ::| |   |:|  |:|  |:  |:| ::|
Db   367 ACNMGNCARCCNGARCAYTTYTYTNYTNCCNGARAYAARCARAAYAARCAR 426

Qy   491 tncayccngayaayaytnathaaraaraaaycentywsngargngnaarttyaarytng 550
         |:|  ::||   |:|  :|:|: |:|: | |:|  :|  |:: |
Db   427 WSNWSNNAARCARWSNCCNCCNACNWSNAARYTNYTNACNGNGMGNGMGNCARACNWSN 486

Qy   551 cngcngarathtgyathtgyaaygargaryttnaaygtnaayccncargayaayggngara 610
         |:|  :|:|  |:|  :|  |:| ::| |  :|  |:|  |:|
Db   487 CCNTGYWSNCAYGGNGTNMGNYTNAARGGNTGGTTWSNMGNCARWSNWSNWSNCCN 546

Qy   611 ayathwsntggacntgycaarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc 670
         |:|  |:| ||  |:|  |   |:|  :|:  |::|  :| :|  |
Db   547 ACNACNTTYTTYACMGNGNGCNGARMGNGAYTTYTGGGCNYTNMGNGGNWSNGGGAR 606

Qy   671 cnmgnmgnnaartggttytgyggnacnggnccnggnwsnsnytntgytgygtncarccmgng 730
         |:|  |   |:|  |:|  ::|  |::|  |:|  |:|  |
Db   607 CARYTNCAYGAYTAYGCNGARMGNGARWSNGGNCAYACNATGMGNGGNCARATHGARGTN 666

Qy   731 ayytngtnccntgytgygtnaaywsngcngtngcnwsngargngngcnwsnccnaarc 790
         |:|  |   |:|  |:|  |   |:|  |   |:|  |
Db   667 ATHCCNTGYAARATHTGYGGNGAYAARWSNWSNGGNATHCAYTAYG---GNGTNATHACN 723

Qy   791 cntggcarytnccnwsngng tngarccngtnggngcnaaataaarwsnmgnathgargtnt 850
         |:|  |   |:|  ::|  |:|  |   |:|  |
Db   724 TGYGARGGNTGYAARGGNTTYTTYMGNGNWSNCARCARAAYAAAYGCNWSNTAYWSNTGY 783

Qy   851 gggarccncconathmgntycaraarathtaygggnaaycentggatgccnmgncaraart 910
         |:|  |   |:|  ::|  |:|  |   |:|  |
Db   784 CCNMGNCARMGNAAAYTGTYTNAATHGYGGNGAYGGNACNAAAYMGNAAAYMGNTGYCARCAYTGYMGN 843

Qy   911 tygcngtnggngtnggnwsnsntggngnacnwsngcnmgntngtncaraargnaayg 970
         |:|  |   |:|  ::|  |:|  |   |:|  |
Db   844 YTNCARAARTGYYTNGCNYTNGGNATGWSNMGNGAYGCNGTNAARTTYGGNMG---NATG 900

Qy   971 tnggntgggarccncncaymgtncnwsnggngcnccnwsnwsnmgngncngtnmgnm 1030
         |:|  |   |:|  ::|  |:|  |   |:|  |
Db   901 WSNAARAARCARMGNGAYWSNYTNAYGCNGARGTNCARAARCAYCARCARMGNYTNCAR 960

Qy  1031 gnwsncncccnwsnsmngnytncaraarggnmgnwsnacngaywsnsnytncarcaygtnc 1090
         |:|  |   |:|  ::|  |:|  |   |:|  |
Db   961 GARCARMGNCARCARCARWSNGGNGARGCNTGNCNGGNGTNTAYWSNWSWSN 1020

Qy  1091 cngaraartwsnacngayacncartgycarccngtnaargcngcnggatggarwsngtnc 1150
         |:|  |   |:|  ::|  |:|  |   |:|  |
Db  1021 ATHWSNAAYGGNYTNWSNAAYYTNAAYAAYGARACNWSNGGNACNTAYGCNAAYGGNCAY 1080

Qy  1151 cntaya 1156
```

```
Db     1081 GTNATH 1086
            |::
RESULT 9
US-08-463-911-6
;  Sequence 6, Application US/08463911
;  Patent No. 5869330
;  GENERAL INFORMATION:
;  APPLICANT: Scherer, Philipp E.
;  APPLICANT: Lodish, Harvey F.
;  TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
;  TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
;  NUMBER OF SEQUENCES: 7
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;  STREET: Two Militia Drive
;  CITY: Lexington
;  STATE: Massachusetts
;  COUNTRY: USA
;  ZIP: 02173
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: Patentin Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/463,911
;  FILING DATE:
;  CLASSIFICATION: 530
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Granahan, Patricia
;  REGISTRATION NUMBER: 32,227
;  REFERENCE/DOCKET NUMBER: WHI95-05
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (617) 861-6240
;  TELEFAX: (617) 861-9540
;  INFORMATION FOR SEQ ID NO: 6:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 1313 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: 73..804
US-08-463-911-6

Query Match     3.8%;  Score 51.8;  DB 2;  Length 1313;
Best Local Similarity  34.2%;  Pred. No. 0.00024;
Matches  75;  Conservative  36;  Mismatches 108;  Indels  0;  Gaps  0;

Qy  158 gycngayathgarmgnwsngncnttyacngtnaarytnwsnggnaarytnccnytnccnt 217
        |:|:|||||:|||||:|:|||||:|||||:||||:|||||:|||  | |||
Db  392 GTGCCTAYGTAYACCGCTCAGCAATTCAGTGTGGGGATTGGAGACTYACGTYACGTTATATCCCCA 451

Qy  218 tyaarccnathathttyacngngtnytntayaayagcncarmgngayytnaargargcna 277
        |:||||||:|||:||:|||||:||:|:|||||:|||  |:|
Db  452 ACATGCCCATTCGCTTTACCAAGATCTTCTACAATCACCAAAACACCATATGATGGCTCCA 511

Qy  278 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnswsnttygaygtngarytnc 337
        ||  |||:|||||::||||:||:|||
Db  512 CTGGTAAATTCCACTGCAACATTCCTGGGCTGTACTACTTTGCCTACCACCATCACAGTCT 571

Qy  338 aycaytgyaargtnaayathtggytnatgmgnaarcara 376
        |:||  ||||||:|||:||||  |:||| |:||  |
Db  572 ATATGAAGGAGGATGTGAAGGTCAGCCTCTTCAAGAAGGACA 610

RESULT 10
US-09-140-804-9
```

```
;  Sequence 9, Application US/09140804
;  Patent No. 6197930
;  GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT: Humes, Jacqueline M.
;  TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
;  FILE REFERENCE: 97-49
;  CURRENT APPLICATION NUMBER: US/09/140,804
;  CURRENT FILING DATE: 1998-08-26
;  EARLIER APPLICATION NUMBER: 60/056,983
;  EARLIER FILING DATE: 1997-08-26
;  NUMBER OF SEQUENCES: 47
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 9
;  LENGTH: 4517
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-140-804-9

Query Match     3.8%;  Score 51.8;  DB 4;  Length 4517;
Best Local Similarity  34.2%;  Pred. No. 0.0008;
Matches  75;  Conservative  36;  Mismatches 108;  Indels  0;  Gaps  0;
US-09-140-804-9

Qy  158 gycngayathgarmgnwsngncnttyacngtnaarytnwsnggnaarytnccnytnccnt 217
        |:|:|||||:||   |:|:|||||:|||   :|:|  :|:|
Db  346 gtgccatgtataccgctcagcattcagtgtgggattggagacttacgttactatcccca 405

Qy  218 tyaarccnathathttyacngngtnytntayaayagcncarmgngayytnaargargcna 277
        |:||||:|||:||:||:|||   ||:|:|||||:|||   |:|
Db  406 acatgcccattcgctttaccaagatcttctacaatcagcaaaacaccactatgatgctcca 465

Qy  278 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnswsnttygaygtngarytnc 337
        ||  |||:|||||::||||:||:|||
Db  466 ctggtaaattccactgcaacattcctgggctgtactactttgcctaccaccatcacagtct 525

Qy  338 aycaytgyaargtnaayathtggytnatgmgnaarcara 376
        |:||  ||||||:|||:||||  |:||| |:||  |
Db  526 atatgaaggaggatgtgaaggtcagcctcttcaagaaggaca 564

RESULT 11
US-09-053-866-3/C
;  Sequence 3, Application US/09053866
;  Patent No. 6111075
;  GENERAL INFORMATION:
;  APPLICANT: Xu, Wenfeng
;  APPLICANT: Presnell, Scott R.
;  APPLICANT: Yee, David P.
;  APPLICANT: Foster, Donald C.
;  TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
;  TITLE OF INVENTION: PAR4 (ZCHEMR2)
;  NUMBER OF SEQUENCES: 12
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: ZymoGenetics, Inc.
;  STREET: 1201 Eastlake Avenue East
;  CITY: Seattle
;  STATE: WA
;  COUNTRY: USA
;  ZIP: 98102
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Diskette
;  COMPUTER: IBM Compatible
;  OPERATING SYSTEM: DOS
;  SOFTWARE: FastSEQ for Windows Version 2.0
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/09/053,866
;  FILING DATE:
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:
;  FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
```

; NAME: Leith, Debra K
; REGISTRATION NUMBER: 32,619
; REFERENCE/DOCKET NUMBER: 98-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6674
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-053-866-3

Qy   302 cnggnaaytaytaywsnwsnttygarytncayctgyaargtnaaayathtggy 361
          |:|:|
Db 1071 YTGRAANARNCCNGNCNCKNACYTTRTCNCKRAAYTCNGCNSWNACRTARTARTADATRA 1012

Qy   362 tnatgmgnaarcarrathytngcnaayaayaargaargarathwsnaarcarwsnathcarg 421
          :|: :
Db 1011 NGGRTCNACRCANSWRTTNARNGTNSWNARNGGNACRTANGCNCNCRTANAR 952

Qy   422 argtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnw 481
          | | |:
Db   951 RTTNCCCANGCNSWNGGNSWNGGRTCNSWRTARTGNARNARNARNARRTTNSWNGG 892

Qy   482 sngaraayytncayccngayaaygtnathaaraaraaayccnttywsnargrggnaart 541
          | |::|
Db   891 NACRAARAANGCNNACNGCNSWNGCNARNACNACNGCNGTNARNCKNARNGCRTCNCCRTA 832

Qy   542 tyaarytngcngcngarathtgyathtgyaaygargargtnaayccncaargaya 601
          | |::
Db   831 NCKNCKNCCNSWNGCNGCNARNGTRTGNARNGTNGCNCRTARCANARNARCATNGCNAR 772

Qy   602 ayggngaraayathwsntggacntgycarmgnqwsnwsncarcarwsnathaarwsnytng 661
          | |:
Db   771 NARNGGNARRARARCANCCNARRARNGCNARRCANGCTRAANGCNCGGYTGCCARTGNSWNGC 712

Qy   662 cntggmgncnmgnmqnnaartggttytytgyyggnacngncnggnwsnytntgytgyytgntc 721
          |:|:|
Db   711 YTGNGCRTCNARNGGNARNGCRTCRTGRCANARNACNNCKRTCNSWNCKNGCCNARNCKRAA 652

Qy   722 arccnmgnqayytngtnccntgyygtnccnqtnaaywsnqcngtngarggngcnw 781
          :| |
Db   651 NGGTYTGNCKYTGNARNGTNARNGGNARNGCNCNARNGCNARNGTNARNGGCCAT 592

Qy   782 snccnaarccntggcaritytnccnwsnggnqtngarccngtngcnaaraarwsnmgna 841
          | |::
Db   591 RCANARNCCNARNGCNARNACNCKNCKNCNCCNGNARRGRTGNACNAR 532

Qy   842 thgargtntgggarccncnaathmgntycaraaraathtaygqnaaycntgqatgccnm 901
          :| |
Db   591 NARNGGNARRARARCANCCNARRARNGCNARRCANGCTRAANGC 532

Qy   902 gncaraarttygcngtnggngtnggqnwsnstggmgnqacnwsnmgngtngtncara 961
          ||:|
Db   471 RTANARNGCNGNCNGTNGCNARNCKRCANGCNCNCRAANGCCANCKYTGNCCNCCNCK 412

Qy   962 arggnaaygtnggqntgggnqarccncnacayngngtnccnwsnwsnwsnmgng 1021
          |:|::
Db   411 NARRTGCRTANGCDATNCKKNGGNGGNARNGCNARNARRTCNGCNGTNGCNAR 352

Qy 1022 cngtnmgnmgnwsncccnccnwsnwsnmgnytncaraargqngmgnwsnacngaywsnytnc 1081
          :|::
Db   351 RTTCATNARNARCATNGTNSWNGGNWSNGGNVARNCKNGGGNGCYTGNGTTNGCNARNACCCANARNGC 292

Qy 1082 arcaygtnccngaraarwsnacngaycancngtnaargcngcnggnatgg 1141

```
RESULT 13
US-09-351-414-3
  Sequence 3, Application US/09351414
; Patent No. 6265199
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT: Baindur, Nand
;  APPLICANT: Deisher, Theresa A.
;  APPLICANT: Bishop, Paul D.
;  TITLE OF INVENTION: DISINTEGRIN HOMOLOG
;  FILE REFERENCE: 98-29
;  CURRENT APPLICATION NUMBER: US/09/351,414
;  CURRENT FILING DATE: 1999-07-09
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 3
;  LENGTH: 2088
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: zdint1 amino acid degenerate sequence
;  FEATURE:
;  NAME/KEY: variation
;  LOCATION: (1)..(2088)
;  OTHER INFORMATION: n is any nucleotide
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)..(2088)
;  OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3

Query Match          3.58;  Score 48.8;  DB 4;  Length 2088;
Best Local Similarity 11.8%;  Pred. No. 0.0026;
Matches 140;  Conservative 286;  Mismatches 749;  Indels 8;  Gaps 3;
```

```
Qy    190 aarytnwsnggnaarytnccnytnccnttyaarccnathathttyacnggngtnaytntay 249
          ::|::: ::|: ::|:|:::|: :|::: |::| :|:|::|: |: ::|:
Db    731 aratgytncaygngartywsnaartaymgncarmgnathaarcacaycngaaycgcngtnc 790
```

```
Qy    716 gygtncarccnmgngayytngcnccntgygtnaawsngngtngcnwsngarg 775
          |:|: |::|: ::: |:|: ||:
Db    676 gayatggngcnytntggcaysnaartaygarwsngayacnytngarcargayytngar 735
```

```
Qy    776 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtngggncnaaaraarw 835
          |::::|: ::: :|:
Db    736 mgnytntycargarytnmgnccnytntaytnaaycncayacntaygtnmgngcn 795
```

```
Qy    836 snmngnathgargtntggargarccnccnathmgnttycararaarathtayggnaaycntgga 895
          :|::|: :|: ::
Db    796 ytncaymgncaytaygnccngaryytngmggnccnathccngcncaytn 855
```

```
Qy    896 tgccnmgncacaaarttycgcngtngggntngngnwsnsntgggnacnwsngcnmgmgtng 955
          ::| |:: :: :|
Db    856 ytnggngaraayacnytngcncarwsntggtnaayathtngayccngtnytnccntty 915
```

```
Qy    956 tncaraargnaaygtnggntggaarccnccnccacaymgngtnccnwsnggngcnccnwsnw 1015
          ::|:|:|: ||:: :|||::: |: |||::
Db    916 ytnaararathccngargaygtnacaaaratnatggtnatgngccnacngaygngangargtn 1095
```

```
Qy   1016 snmngncngtnmgmgnwsnccnccnwsnwsnmngnytncaraargngngmgnwsnacngayw 1075
          :: |: ||:|: :|: |::
Db    976 aarytnatgytngaargarcgnaracnttyyacntaayytngngnytngcnytncncncn 1035
```

```
Qy   1076 snytncarcaygtncngaraarwsnaacngayacncaartgycarccngtnaargcgncng 1135
          :: ||: :|||: :|:| :|
Db    976 snccnccnwsnttyggaararaarytnatgytnatgngngccnacnacngaygggnmgnaargtn 1095
```

```
Qy   1136 gnatg 1140
          :||:
Db   1096 gartg 1100
```

```
Qy    250 aaygcncarmgngayytnaargargcnatggnggntnttygcntggymgngtnccnggnay 309
          :||:: :|: |||: |:|: ::::
Db    791 ayyttnathwsnmngtnacnttycaytayaamgnwsnwsnytnwsntayttygarggng 850
```

```
Qy    310 taytaywsnwsnttygaygtngarccngtnggngcnaaaraarwsnmgnathgargtn 369
          ::|:: :::|: ::|: :|||:|
Db    851 tntgywsnmngnacnmgngngtnggngtngarytncaycatntcgyaargtnaaytncatgmgn 910
```

```
Qy    370 aarcarathytngcnaaygaragarathtaywsnaarcacaaythcaargargtnacn 429
          |::: ::: : |::|: |: |
Db    911 argtngnytnwsncarwsnytngcncaraayytngnathcaargargcaarwsnsnmgna 970
```

```
Qy    430 tgggtnytnaytnaargcnttywsnttyatymtmgngargcngrncaayaarwsnsngaray 489
          ::| :|:|| |: ||: :|:::|::
Db    971 arccnaartgygaytgyacngarwsntgggnngtygytnathgargayaaarwsnga 1030
```

```
Qy    490 ytncayccngayaaygtnathaaraararaaycnttywsngargnaarttyaarytn 549
          |:|| ||: : :||::|||
Db   1031 sncaywsnmgnaarttywsnaartgywsnathytngartaymgngayttyytncarmgng 1090
```

```
Qy    550 gcngcagarathtgyathtgyaaygargarytnaaytnaaycncngaryaaygngar 609
          :|: :: |:|::||: |:: ::
Db   1091 gnggngngncntgyytnttyaaymgnccnacnaaryttnttygarccnacnacngaygggna 1150
```

```
Qy    610 aayathwsntggacntgycarmgnwsnwsncarwsnathaarwsnytncngcnacngarytngcng 669
          ::|:||:: ::||:||: |::|:
Db   1151 ayggntaygtnga--rgcnggngargarytgywsnytnycaygtngaartgytaygg 1208
```

```
Qy    670 ccnmgnmgnaaartggttytgyggnacnggnccnggnwsnytntgytygytncarccnmgn 729
          |::|: |::|:::||
Db   1209 nytntgytgyaaaaartgywsnytnwsnaaygngngcncaytgywsng-ayggnccntgyt 1267
```

```
Qy    730 gayytngtnccntgygtnccngtnaaywsngngcngtngcnwsngargngcnwsnccnaar 789
          ::::|
Db   1268 gyaayaayacnwsntgytytnttycarccnmggtnggtaygartgymgngaycngtnaayg 1327
```

```
Qy    790 ccntggcarytnccnwsnggngtngarccngtnggngcnaaaraarwsnmgnathgargtn 849
          :||: |
Db   1328 aartgygayathacngartaytgyacngngaaygwsnggngcaartgycncnaaayytncaya 1387
```

```
Qy    850 tgggarccnccnathmgnttycararaarathtayggnaaycntggatgcnmgncaraar 909
          :|:|: |::|
Db   1388 arcargaygggntaygcntgyaaaycaargcngmgntgytayaaaggngaartgyaara 1447
```

```
Qy    910 ttygcngtnggngtnggnwsntggmgnacnwsncnmgntngtngtnccaraargnaay 969
          :|::: |:
Db   1448 cnmgngayaaycaartgycartayathtgggngnacnaaargcngcnggnwsngayaarttyt 1507
```

```
Qy    970 gtnggntgggarccnccncaymgngtnccnwsngngcnccnwsnsnmgngcntngmgn 1029
          :| :: |: |
Db   1508 gytaygaraarytnaayacngaragaraarggna----aytngyggnaarggaygg 1562
```

```
Qy   1030 mgnwsnccnccnwsnwsnmgnytncaaraarggngmgnwsnacngaywsnsnytncacaygtn 1089
          :|:
Db   1563 ngaymgntggathcaartgywsnaarcacayaaygtntttyggnttyytntytgyacnaa 1622
```

```
Qy   1090 ccngaraarwsnacngayacncaartgycarccngtnaargcngngngnatggaarwsngtn 1149
          :|:: |: | :|::|
Db   1623 yytnacnmgngccnmgnathggncarytncargngnarathathccnacnwsnttyta 1682
```

```
Qy   1150 ccntayaaacngtngcngarytnacnaaaracngtnggnnathtayytnaytncaytgy 1209
          :|: ||: :|: |: :
Db   1683 ycaycaarggnmgntnathgaytgywsnggngncncaytgtnrytngaytgaygaycncga 1742
```

```
Qy   1210 caygayytngaygtmgncaygngtnaarmgngaycaytgggncnytnmgnttygay 1269
          :|: ||:::|:|
Db   1743 ygtngntayytngargaygganacnccntgyggccnwsnatatggtggytntgnagnaa 1802
```

```
Qy   1270 tgyccnacngngnttygmgnacntayatgggnccngtnccnytntgytyggncaarttytty 1329
          :|: ::
Db   1803 rtgyytncaratncaatgcnytnaaytagtgwsnsntgycncytngaywsnaarggnaaragt 1862
```

```
Qy   1330 ccnttyggnacngcngtnttttyaccnacaartgyytntayytncayt 1372
```

Db   1863 ntgywsnggncaygyngtntgywsnaaygaargcnacntgyath 1905

RESULT 14
US-09-007-005-17
Sequence 17, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihe
TITLE OF INVENTION: FUSIONS
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE REFERENCE: 0078G/350003
CURRENT APPLICATION NUMBER: US/09/007,005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 1997-11-06
EARLIER FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(289)
OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match   3.5%; Score 48.6; DB 4; Length 289;
Best Local Similarity   10.1%; Pred. No. 0.00044;
Matches   27; Conservative 101; Mismatches 139; Indels   0; Gaps   0;

Qy   389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnaaargcnt 448
          ::: |::::::::: :|||::||::: |::: |::::||
Db   19 rarcrurarururarcrararurarurgnrnrsrnrnrsrnrnrsrnrn 78

Qy   449 tywsnttyathmgnargncngarcayaaarwsnwsngaraayytncayccngayaaygtna 508
          ::: | |::: | | |::: |::: | | |
Db   79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy   509 thaaraaraaraayccnttywsngargnaarttyaarytngcngngarathtgyatht 568
          ::: | |::: |::: |::: ::::
Db   139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy   569 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 628
          ::: |::: | ||::::::::: :
Db   199 rsrnrnrsrnrnrsrnrnrsrcrargrcrgrurarcrcrurcrururgrcrgrc 258

Qy   629 armgnwsnwsncarcarwsnathaarw 655
          |: |::: |::::
Db   259 ruaaaaaaaaaaaaaaaaaaaaaaaa 285

RESULT 15
US-09-244-796-17
Sequence 17, Application US/09244796
Patent No. 6281344
GENERAL INFORMATION:
APPLICANT: Szostak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihe
TITLE OF INVENTION: FUSIONS
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE REFERENCE: 0078G/350007
CURRENT APPLICATION NUMBER: US/09/244,796

CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 1997-11-06
EARLIER FILING DATE: 60/064,491
EARLIER APPLICATION NUMBER: 09/007,005
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 09/007,005
EARLIER FILING DATE: 1998-01-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(289)
OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match   3.5%; Score 48.6; DB 4; Length 289;
Best Local Similarity   10.1%; Pred. No. 0.00044;
Matches   27; Conservative 101; Mismatches 139; Indels   0; Gaps   0;

Qy   389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnaaargcnt 448
          ::: |::::::::: :|||::||::: |::: |::::||
Db   19 rarcrurarururarcrararurarurgnrnrsrnrnrsrnrnrsrnrn 78

Qy   449 tywsnttyathmgnargncngarcayaaarwsnwsngaraayytncayccngayaaygtna 508
          ::: | |::: | | |::: |::: | | |
Db   79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy   509 thaaraaraaraayccnttywsngargnaarttyaarytngcngngarathtgyatht 568
          ::: | |::: |::: |::: ::::
Db   139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy   569 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 628
          ::: |::: | ||::::::::: :
Db   199 rsrnrnrsrnrnrsrnrnrsrcrargrcrgrurarcrcrurcrururgrcrgrc 258

Qy   629 armgnwsnwsncarcarwsnathaarw 655
          |: |::: |::::
Db   259 ruaaaaaaaaaaaaaaaaaaaaaaaa 285

Search completed: July 8, 2002, 16:20:40
Job time: 12164 sec

```
OM nucleic - nucleic search, using sw model
                                    GenCore version 4.5
                                    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:           July 8, 2002, 13:28:40 ; Search time 3393.27 seconds
                                    (without alignments)
                                    5477.108 Million cell updates/sec

Title:            US-09-997-610-3
Perfect score:    1377
Sequence:         1 athgtngtnathccngtnyt.........gyytntayytncaytgyyatg 1377

Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0
Searched:         13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters:    27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        EST:*
                  1: em_estba:*
                  2: em_esthum:*
                  3: em_estin:*
                  4: em_estnu:*
                  5: em_estov:*
                  6: em_estpl:*
                  7: em_estov:*
                  8: em_estro:*
                  9: em_htc:*
                 10: gb_est1:*
                 11: gb_est2:*
                 12: gb_gss:*
                 13: gb_htc:*
                 14: em_gss_hum:*
                 15: em_gss_inv:*
                 16: em_gss_pln:*
                 17: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	479.2	34.8	2615	10	BE420422	BE420422 32-393 hu
2	450.2	32.7	2009	11	BC004496	BC004496 32-1514 h
3	398	28.9	1514	10	BE512633	BE512633 32-1514 h
4	336.8	24.5	870	10	BI488505	BI488505 603021014
5	335.6	24.4	1050	10	BM472108	BM472108 AGENCOURT
6	321.8	23.4	1036	10	BM471183	BM471183 AGENCOURT
7	318.6	23.1	898	10	BM457166	BM457166 AGENCOURT
8	313	22.7	717	12	AG086951	AG086951 Pan trogl
9	306.6	22.3	736	10	BI754555	BI754555 603023384
10 c	300.4	21.8	654	12	AG112666	AG112666 Pan trogl
11	296	21.5	676	12	AG064424	AG064424 Pan trogl
12	287.4	20.9	691	12	AG062447	AG062447 Pan trogl
13	286.2	20.8	720	12	AG116938	AG116938 Pan trogl
14	283.8	20.6	623	12	AQ037711	AQ037711 CIT-HSP-2
15	283.2	20.6	1036	12	AG097258	AG097258 Pan trogl
16	282.2	20.5	704	12	AQ588828	AQ588828 CITBI-El-
17	281.6	20.5	659	12	AG092858	AG092858 Pan trogl
18	281.4	20.4	701	12	AQ588873	AQ588873 CITBI-El-
19	281	20.4	729	10	BM045873	BM045873 603624875
20	279.6	20.3	760	10	BG281182	BG281182 602402107
21	279.4	20.3	697	10	AG018926	AG018926 Homo sapi
22	278.8	20.3	885	10	BC761979	BC761979 602718930
23	277.2	20.1	1182	10	BM451317	BM451317 AGENCOURT
24	275.2	20.1	639	12	AG115932	AG115932 Pan trogl
25	275	20.0	676	10	AG125093	AG125093 Pan trogl
26	274.6	20.0	721	10	BE728616	BE728616 601563827
27	274.4	19.9	671	12	AG045131	AG045131 Pan trogl
28	272.2	19.8	877	10	BI916691	BI916691 603178655
29	271.4	19.7	704	12	AG070563	AG070563 Pan trogl
30	271.2	19.7	686	12	AQ020199	AQ020199 CIT-HSP-2
31	268.8	19.5	660	12	AG119684	AG119684 Pan trogl
32	268.6	19.5	746	10	AG057341	AG057341 Homo sapi
33	268	19.5	623	10	AG001408	AG001408 Homo sapi
34 c	266	19.3	699	9	AG075663	AG075663 Pan trogl
35	266	19.3	699	9	AL602279	AL602279 DKFZp686B
36	264.8	19.2	543	12	B94403	B94403 CIT-HSP-217
37	263.8	19.2	675	12	AG096655	AG096655 Pan trogl
38	261.4	19.1	833	9	AU140358	AU140358 Homo sapi
39	261	19.0	771	12	AG001229	AG001229 Homo sapi
40	260.6	19.0	716	12	AG104336	AG104336 Pan trogl
41 c	260.6	18.9	940	10	BG679573	BG679573 60262761
42	260.6	18.9	955	10	BG681769	BG681769 60262787
43	257.4	18.7	649	10	AW857140	AW857140 RC1-CT030
44	256.6	18.6	900	10	BM449699	BM449699 AGENCOURT
45	255.4	18.5	676	12	AQ239408	AQ239408 CIT-HSP-2

ALIGNMENTS

```
RESULT    1
BE420422
LOCUS       BE420422               2615 bp    mRNA    linear    EST 11-OCT-2001
DEFINITION  BE420422 32-393 human bone marrow cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BE420422
VERSION     BE420422.1    GI:16041640
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2615)
  AUTHORS   Lu,X., Cui,L. and Li,Y.
  TITLE     DDRT-PCR from B cell
  JOURNAL   Unpublished (2000)
COMMENT     Contact: xingwu Lu,liangxian Cui,yonghai Li
            Department of Biochemistry
            Institute of Basic Medical Science, Peking Union Medical College
            DongDan SanTiao 5, Beijing, P.R.C, 100005
            Tel: 86-010-65296951
            Email: luxingwu@263.net
FEATURES             Location/Qualifiers
     source          1..2615
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="human bone marrow cDNA library"
                     /tissue_type="bone marrow"
BASE COUNT      705 a    531 c    641 g    738 t
ORIGIN

Query Match      34.8%;  Score 479.2; DB 10;  Length 2615;
Best Local Similarity   47.1%;  Pred. No. 4.6e-98;
Matches  500; Conservative  222; Mismatches  299; Indels  41; Gaps  4;

Qy    352 aayathtggytnatgmgnaarcaarathytngcnaayaaryaargaratthwsnaarcarcar 411
          |||:|||| | |  |:| :| ||||  ||| || |  |   ||||:|||:|||::|||:|||:
Db    494 AACCTTGAGCTTGAAGAAGAGATCATTTAGGTGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 553
```

```
LOCUS       BC004496                2009 bp    mRNA    linear   HTC 12-JUL-2001
RESULT  2
BC004496
DEFINITION  Homo sapiens, Similar to hypothetical protein FLJ14058, clone
            IMAGE:3831313, mRNA.
ACCESSION   BC004496
VERSION     BC004496.1  GI:14709139
KEYWORDS    HTC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dicknhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan
            This clone has the following problem: frame shifted.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 14 Row: d Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
FEATURES             Location/Qualifiers
     source          1..2009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3831313"
                     /tissue_type="skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_20"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      556 a    429 c    513 g    511 t
ORIGIN
```

```
Query Match             32.7%;  Score 450.2;  DB 11;  Length 2009;
Best Local Similarity   45.2%;  Pred. No. 1.6e-91;
Matches    452;  Conservative  214;  Mismatches  321;  Indels  13;  Gaps  3;
```

```
Qy    412 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 471
          ::|:|:|:||:|.||::||:  |||:::|||||:..|:||::::: |||:|::.|| :::
Db    554 AGCATTCAAGAGAGTHAACTHGGGTHCGGTCGTTGTTAAAGGCATTTTAGTTTTATAAGGGAAGCAGG 613

Qy    472 cayaayaarwsnwsngararaayytncaycccnqayayaygtnathaararaaraayycnttywsn 531
          ::|:::.||  ::||||::|:|::||||||:|||::::|||:.||||::||:::||||::::||:
Db    614 CATGAAAGTTTGGAAAATTTGCAGCCTGACAATCAAATAGAAAAGAAAATCCCATATTCT 673

Qy    532 gargqnaarttyaarytngcngcngararathtgyaaygargaryytnaaygtnaay 591
          |:|: ||: ::|||:|||:||:|||:|:||||::||::||||:|:|: |||.:|:|: ||:
Db    674 GAGAGAAAATCAAGCTGCTGCAGAGAATTTGCAYCAYTATGTTAAYACAAGGAGYTAAAGTTAAT 733

Qy    592 cncaargayaaygngaraayathwsntggacntgycarmgnnwsnmgncarcarwsnath 651
          |: ||::.| |:|||:::||||::||||:|||::||||||:|||:::||||||::||:|::|
Db    734 CCCAAGACAAATGGGAAAAATGYTCCAGGCATGCCAGAGYCTTCACAGCAGCCCCTC 793

Qy    652 aarwsnytytngggmgnccnmgnmg--naartggttgtytgygnacnggncnggnwsn 708
          :::||:|::|||||::||||:||||   :::|||:.||:|:|:|||||:::|||||||:
Db    794 CCATCACAGGCCTCGCAGGCCTAGGAGGAAAAAGTGAATCYTCTTWGGCCAGCCCAGGGTCC 853

Qy    709 ytntgyttgygttncarccnmgnqayytngtgtnccntgygtnaccngtnaaywsngc------ 761
          |||:||||:|:|||::||||:||:||||:||:||::||:|||::|::||||||::|:
Db    854 CCAYGCTGYTGTGCAGCCTGGYGACTTGGTGTCCCTYGTGTCCCAGCTGCCTCCCAGCCATGGGCT 913

Qy    762 ----------ngtngcnwsngargngngncnwsncncnaarccntgg 795
          .:| |:||:||:||:|||||||:::||::||::|||||||||:|||
Db    914 GAAAAGGGCCCAACATAGAGCTCCGGATTGTGGCTTCAACAGAGGGGTCGCAGCGTGTAAGCCTTGG 973

Qy    796 carytnccnwsngngngtngaccnmgngcnaaaarwsnmgnathgargtntgggar 855
          |||:|:||:||:||::|::||:||||:|:|:||||:||||:|||:::|||:|:||||
Db    974 CAGCTTCCAYGTGGTGTGGAGCCTCCCAGTGCACAGAAGTCAAAAATTGGGGTTTGGGAA 1033

Qy    856 cnccnaathmgnttcycaraarathtayggnaayccntggatgcnmgncaraarttygcn 915
          .:| |:||:||||||::||||:||::||:||::||::||:|||:||:|:||||:||:||
Db    1034 CCCTCGCCTAGATTTCAGAAGACGTATGGAAAATGCCTGCCGCAGAGCGTGTAAGCCTGCT 1093

Qy    916 gtnggngtnggnwsnwsntggmgnacnwsngcnmgntngtngtncaraargngnaaygtnggn 975
          |:|||:|::||:|||||||||||:|:||||:||:|:::||::||:||||||:|:|||::|||:|
Db    1094 GTAGGGAGGGACGCCCTCAYGGAGAATCTCTGCTGYCAGGGACGTACAGAAAGGGAAATGTGGGA 1153

Qy    976 tggaarcncnccncaaymgntgtncnwsnwsnngngcngtnmgnmgngnwsn 1035
          |||||:|::|:||:|||::||::|||:::||||||||:||||:||:|:||||
Db    1154 TCGGAGCCCCGCAGCAGCAGAATCCCTACTGGGGCACCGCGCCTGTGAGAGAGGG 1213

Qy    1036 cnccnaathmgnttcycaraaraggnngnmgnaaywsnytncaraygtnccngar 1095
           .:| |||:||||||||::||||:|||:||:||:||||::||||||:||:||:|||
Db    1214 CCACCATCTCCAGACCCCAGAAGATAGATCCACTGACAGCGTTGCTCATGCACCTGA 1273

Qy    1096 aarwsnnacngaycarcarcngtnaaargcncnggnatggarwsngtnccntay 1155
           :::||::||:||:|||||:||::|:||||::||||:||||||:||:||:||:|||
Db    1274 AAAACCGCAGA-----CAACACCACCCATGAAAACGCAGGAGGGCGTACCCTGC 1329

Qy    1156 aaracngtngngtnacacnngayntnacacngtnggnathtayytnytnhcaytgycaygay 1215
           ::|:||:||:||:||:||::||:||:||:||:|||||||:|||:||:|:|:|||:||:|||:|
Db    1330 AAAGCCACAGGGCGGCGAGCTGCCCAAGATCATCAGGGAACCACCTCTTGCATCAGCATGAC 1389

Qy    1216 ytngaygtngmgncaygghgntnaarmgnnaygcaytytygayytngaytgyccn 1275
           ::||||||||:||:||:||  |::|||:|||:||:::|:|||:::||:||:||||:
Db    1390 CTGGAYGTGTGAGATTTGGAGTTTAAGGAAGCAGCATAAAAGTTCAGAAAYTTGCAGCCTGACAAATGTGAT 1449

Qy    1276 acnggnttymgnacntayatgggngccngt-nccnytntgytyggncaarttyyttyyccntt 1334
           ||:|||:|:||:||:|||||||||:||:||  :||:|:|:||:|||:||:|||::|::||::||
Db    1450 TCTAGATTTCAGAGTTCAGCTGACTTGCAYGGGCGTGCCCTTGTTTACTCAAYTTCTCCCATT 1509

Qy    1335 yggnacngcngtntttyacnacartaygtdgyat 1376
           :|||||:||:||:||::||:||:||:||:||.|||
Db    1510 TGGAACAGCTGTGTTTACAACAATACCTGTACCCCATTGTAT 1551
```

```
Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          |:|||:|||::||::||:||:||::||||||||||:||:||||||:|::|:||:|||:||
Db    489 AGAAGAAAATTTCTATCCAGCAAAGCAHTTCAAGAGAGTTGACTTGGGTGCCTGTTAAAGGCATT 548

Qy    450 ywsnttyathmgngargcngarcayaayaarwsnwsngararaayytncaycccnqayayaygtnat 509
          ::|::||:||:|||:|:||:||:|::::|::||||||::|:|::||:||||:|||:|||::||:||
Db    549 CAGATTTATAAGGGAAGCAGCAGCATAAAAGTTCAGAAAGTTGCAGCCTGACAATGTGAT 608

Qy    510 haaraaaraaaayccnttywsngarggngnaarttyaarytngcngcngararathtgyathg 569
          :::||::|:|::||||:|:||:|||:||:|||::|:|||:|||:||:|||:|:||||:|||||
Db    609 AGAGAGAAAATCAAGCTGCTGCAGAGAATTTGCAYCAYTATGTTAAYACAAGGAGYTAAAG 668

Qy    570 yaaygargaryytnaaygtnaayccncaargayaaygngaraayathwsntggacntgyca 629
          :::|::|||:|:|: |||.:|:|::||::.| |:|||:::||||::||||:|||::||||||
Db    669 TAACAAGAACGCCGCAYGTTAATCCCCAAGACAAATGGGAAAAATGYTCCAGGCATGCCAGAGYCA 728

Qy    630 rmgnnwsnwsncarcarwsnathaarwsnytytgncntggmgnccnmgnmgngnaartggttytg 689
          ::| |:||:::||:||::||:|||:::||:|::|||::||||::||||:||||::|:|||::|::|
Db    729 GAGGTCTTCACAGCAGCCCCTCCCAGYCACAGCAGCCTAGGAGGAAAAAGTGAATCYTCTTWGG 788

Qy    690 yggnacngcncnggnwsnsnytnt------gytgyggtncarccnmgnqayytngtgtnccn--- 740
          :|||||:||:||:|||||::||:||       |:|:||:||::||||:||:||||:||::|||
Db    789 GCCAGCCCAGGGTCCCAYGCTGYTGTGCAGCCTGGYGACTTGGTGTCCC 740
```

```
RESULT  3
LOCUS       BE512633                1514 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  32-1514 human B lymphocyte cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BE512633
VERSION     BE512633.1  GI:16041645
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Lu,X., Cui,L. and Li,Y.
  TITLE     DDRT-PCR from B cell
  JOURNAL   Unpublished (2000)
COMMENT     Contact: xingwu Lu,liangxian Cui,yonghai Li
            Department of Biochemistry
            Institute of Basic Medical Science, Peking Union Medical College
            DongDan SanTiao 5, Beijing, P.R.C, 100005
            Tel: 86-010-65296951
            Email: luxingwu263.net
            full cDNA sequence.
FEATURES             Location/Qualifiers
     source          1..1514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="human B lymphocyte cDNA library"
                     /tissue_type="bone marrow"
BASE COUNT      380 a    341 c    398 g    395 t
ORIGIN
```

```
Query Match    28.9%;  Score 398;  DB 10;  Length 1514;
Best Local Similarity   47.1%;  Pred No. 1 1e-79;
Matches  427;  Conservative 182;  Mismatches 256;  Indels 41;  Gaps 4;
```

```
Qy    508 athaaraaaraaaaycccntywsngargngnaartyaarytngcngcngarathtgyath 567
          ||||.|||::| :|: | ||.|| ||.| :::.| ||| |||.|| ||.||| ||:||
Db      1 ATAGAAAAGAAAAATCCCATATATTCTGAGGAGAAAAATCAAGCTTGGCTCCAGAAATTTGCATA 60

Qy    568 tgyaaygargaarytnaaaycncacgayaaygngaraayathwsntggactgy 627
          | :||.|||:| ||:| :|||||:| :|| :::| ||| | :| ||| |:|||
Db     61 TGTAACAAGGAGTTAAATGTTAAATCCCAAGACAATGGGAAAAGTCTTCCAGGGCATGC 120

Qy    628 carmgnswsncarcarwsnathaarwsnytngcntggmgnccnmgnmg--naartgg 684
          ||::| |||:||:::| ||.||.| ::: |||::.| || |::| ||||:
Db    121 CAGAGGTCTTCACAGAGCCCCTCCAGGCTGCCTAGGAGCCAAATATATCTGCAA 180

Qy    685 ttytgyggnacnggngccncngghwsnytntgytgyygtncarccnmngayayytngtnccntgy 744
          |:| :.|  |||  |||   |||  || :|:|: :||:| ||:||  ||| |.:| || |::|
Db    181 TCTCTTGGGCAGGCGCCAGGGGTCCCCATGCGTGTGTCCAGCCGTTGGTCCGTTGT 240

Qy    745 gtnccngtnaaywsngc------------------------------ngtngcnwsn 771
          |::| ||:|||:|||| :                            :: |.:
Db    241 GTCCCAGCTGCTTCACGCTGAAAAGGGGCCAACATAGGAGCTCGGATTGTGGCTTCA 300

Qy    772 gargngngcnwsccnaaarccntggcarytnccnwsngngtngarccngtnggngcnaar 831
          || |.|:|| |:| |||.|| |||:|:| || :.| | |:|.| |||:|:|| |:|
Db    301 GAGGGTGCAACGTCAAGCTGTAGGCCTTGGCAGCGTGGTGTTGCACAGCCAGGAAAAAGTGA 360

Qy    832 aarwsnmgnathgargtntggacngyccnac 891
          ||| |||:|.|:| || :||.| | ||||:|
Db    361 AAGTCAAAAAAATTGGGGTTTGGGAACCTCTGCCTAGATTCAGATGTAATGGGAAATGCC 420

Qy    892 tggatgccnmgncaaraarttygcngtngggntyggngtnggngnacnwsngcnmgn 951
          |||||.|  .||||||:||||.||.| |::||| | |  || | ||||| ||.||
Db    421 TGGATGCCCTGGCAGGAAGGTCTGCTGTAGGGGAGGGGGGCCCTCATGGAGAATCTCCTAGG 480

Qy    952 gtngtncaaraggtnggntyggggaarccncacaymgngtnccnwsnggngncccn 1011
          |:| |||||.|| ||::.|  |||| | ||||||||||| | |::| ||.|:|
Db    481 GCAGTACAGAAGGGAAAATGGGAGGCCCCAGCACACAGAAATCCCTACTGGGGCACCG 540

Qy   1012 wsnwsnmgncngtnmgnmgnwsnccncnwsnmgnytncaraaargggmgnwsnacn 1071
          |||||:||.|:|:|| |.||| |||| :|||::||:|||:|:||::.|::::| |
Db    541 CCTAGTGGAGCTGTGAGAAGAGCCACCATCCTCAGACGCCACAGAAATCATCATGGGA 600

Qy   1072 gaywsnytncarcaygtcnngargaraarwsnacngayacnccngagca 1131
          || |||||:||:||:| | :.| ||| | |||.| |.||| ||:|.|||
Db    601 GACAGCTTGCTCAGCAAGCTTGGAAACCTGCACA---CAACACAGCCCCATGAAAGCA 656

Qy   1132 gcnggnatggargwsngtnccntayaaracngtngtncgngarytacacngtngn 1191
          || ||.||.||| ||.|: ||:|||:|| || ||:|.|  || ||.||||:| |
Db    657 GCTAGGCAGGAGGCTGTACCCTGCAAAGCCACGCGGAGAGCTCGCAGATCACTGGA 716

Qy   1192 athatayytnytncaygycaygayytngaygtnmgncayggngtnaaarmgngaycaytty 1251
          |:|||:||.|:||:||||.| |.|| || ||| .||||.|::| || |.| ||||::|
Db    717 ACCCACCTCTTCATCAACATGAGCATGACAGTGTGAGGTTTAAAGGAGAATCATTTT 776

Qy   1252 ggngcnytnmgnttygaytgyccnacnggghttymgnacntayatgggngncngt-nccnyt 1310
          |||:| || |.||.| |  || | |  |:||||::||||||:|:|||.|.| || :| ||
Db    777 GGAGCTTTAAGATTTCAGCTTGACTGCCCTCTAGAGTTTGACTGCATGGGTGCGCCCT 836

Qy   1311 ntgyttyggncarttyttycacntygghnacncngcngtnttnttacacngcntca 1370
          |::||||  || || |||:||: |.| || | ||.|:| ||:|:||||||:||
Db    837 TTGTTTTACTCAATTTCTCCCAATTGGAACAGCTGTGTTTACACAATACCTGTACCCCCA 896

Qy   1371 ytgyat 1376
          |:|:||
Db    897 CTGCAT 902
```

```
RESULT  4

Db   897  TTGTAT  902
          ::|:||
Qy        :::|:||
```

```
LOCUS        BI488505              870 bp    mRNA    linear    EST 28-AUG-2001
DEFINITION   603021014F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191803 5',
             mRNA sequence.
ACCESSION    BI488505
VERSION      BI488505.1  GI:15327733
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 870)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LIAM11479 row: k column: 04
             High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1..870
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5191803"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning). Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      222 a    211 c    256 g    181 t
ORIGIN
```

```
Query Match   24.5%;  Score 336.8;  DB 10;  Length 870;
Best Local Similarity   44.6%;  Pred. No. 7.2e-66;
Matches  378;  Conservative 161;  Mismatches 273;  Indels 35;  Gaps 3;
```

```
Db   242 CTTGTGTGCCCTGTGTCCCACTGCGTCCGCTTCCACGCTGTGACTGAAAGGGGCCACGTAGAGCTCG 301
Qy       |||:||   ||:||| ||:.| |||||:.||| | ||:|: |||||:||| |:|| ||| :|||
Qy   302 GGCIGTGGCCTTCAGANGGCTGGAAGGCTCTAAGCTTTGGCAGCTTCCACGTGTTGACC 361
Db       || |:|| ||  |.|   || :  ||: ||| ||| :|| || || |.||  | |||
Db   362 TACAGGTGCACAGAAGTCAAGAAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGAT 421

Qy   879 htayggnaayccntggatgccnmgncacaraarttcygcngctnggngtnggnwswsntgggg 938
Db   422 GTTTGGAAAAGCCTGGATGCCCAGCCAAAAAGTTTGCTGCACGGAGTGGGGCCCTTATGGAG 481

Qy   939 nacnwsngcnmgngtngtncaaraarggnaaaygtngggarccnccnccaymgngtncc 998
Db       | .:|:||    |   |    |:||| | |||  |: |: |:||
Db   482 AACCTTCTGCTGGGAGCAGTGCAGAAGGAAAAATGTGGGATCGGAGGCCCCGACAGAGAGTCCA 541

Qy   999 nwsngngngcnccnwsnwsngmgncngtnmgngmgwsncnccnccnwswsngngnytncaraa 1058
Db       :: |:| | |||: |::    ||| | |: |:| :  |: |||| || ::| :|
Db   542 TACTGGGGCACTGCCTTAGTGGAACTGTGAGAAGACAACACCGTCTTCAGACCCCAGAA 601

Qy  1059 rgggmgnwsnacngaywsnytncacaygtnccngararaarwsnacngayacncartgyca 1118
Db       | ||:|   |  || |   |:|  | | ||| |:  | | | | |:| |: | |||
Db   602 TGGTAGATCCACCACAGCTTTGGCACCATGTCCGAAAAAGATGCAGGTACTCAATGCCA 661

Qy  1119 rccngtnaargcngcngngganatggarwsngtnccntayaaracncngtngtncgnarytnac 1178
Db       :|| ||||| |:|  ||:|   ||:| |: |:|| || | |:| ||| || :|
Db   662 GCCCATGAAGGCAGCCAGGAGGAGGCGTGRACCCCGCAAAAGCCACCAGGGAGGAGGAGCTGTC 721

Qy  1179 naaracngtngngnathtayytnaytncaytgycayngayytngaygtnmgncaygggngtnaa 1238
Db       ||:||   |: | :| || :|     : | : |: | :| | |:||| ||||:|| |||
Db   722 CAAGACCATGGGAACCCACTCTTGCATCAGCGGTGGATGCCGAGACCTGGACTCAA 781

Qy  1239 rmgngaycayttygngngcnyt-nmgnttygaytgyccnacnggnttymgnacntayatgg 1297
Db       |:||| | ||||| |||||:||  | ::|| |||| | || | |||:| |:||| |||
Db   782 AGGAGGATCATTTTGGAGCTTTAACAAATGTGACCTGCCTGCTGATTTCAGACTTGCATTG 841

Qy  1298 gnccngt 1304
Db       | |||||
Db   842 GGGCCCT 848
```

```
RESULT  5

LOCUS        BM472108             1050 bp    mRNA    linear    EST 05-FEB-2002
DEFINITION   AGENCOURT_6465359 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5539381
             5', mRNA sequence.
ACCESSION    BM472108
VERSION      BM472108.1  GI:18521150
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1050)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LIAM12233 row: m column: 14
             High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1..1050
                     /organism="homo sapiens"
```

```
Qy   493 cayccngayaaygtnathaaraaraarayccnttywsngargngnaarttyaarytngcn 552
Db       |||:||||:||||||||||||||||||| ||  | |||||||||||| |||| |||
Db     2 CAGCCTGATGATGAAGTAGAAAAAAAACCCATTTTTTTTCAGGATAAATTCAAGCCAGCT 61

Qy   553 gcngarathgyathgyaaygargarytnaaygtnaaycyncargayayyggngaraay 612
Db       | |||||| | ||: | |:|| |:|| |:|  | |:||  ||||  |:|||||||
Db    62 GCAGAAATTGCATATGTAGCAAGGAGTCTAATGTTAATCTCCAAGACCATGGGAGAAAT 121

Qy   672 nmgnmgnaartggttytgyggnacnggncnnggncnggnngna wsngtcncarccnmgnga 731
Db       |:|  |:|||| ||||| ||| ::  |  |::  |:  | | | |  |: | ||:||| |
Db   122 GTCTCCGGGCCCATGGTCAGAGACCTTCATGCCAGCCCCTCCCATCACAGGCTGGAGCCAG 181

Qy   672 athwsntggacntgycarmgnwswsncarcaarwsnathaarwsnytntgytgytncarccnmgnc-c 671
Db       |.||| |||| |||  |||  |:|  |  |  :|   |  |: ||:|  |||| |||||
Db   182 GAGGAAAAAGAGGTTCGTGGGGCCCAGGCCATGGGAGGTGTTGTCAGCCTAGGA 241

Qy   732 yytngtnccntgyyggtnccngt---------------------------------naayws 758
Db       :|| ||||| | |||||  | ||||              ||  |||| || |||
```

```
RESULT  6
BM471183
LOCUS       BM471183               1036 bp    mRNA    linear   EST 05-FEB-2002
```

```
            /db_xref="taxon:9606"
            /clone="IMAGE:5539381"
            /tissue_type="melanotic melanoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 2 kb. Library constructed by Life
            Technologies."
BASE COUNT    295 a    262 c    272 g    221 t
ORIGIN

Query Match      24.4%;  Score 335.6;  DB 10;  Length 1050;
Best Local Similarity   43.8%;  Pred. No. 1.5e-65;
Matches 335; Conservative 169; Mismatches 248; Indels 12; Gaps 2;

Qy    390 rgrgaratthwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :|:||.||| ||.|||.|| ||||||||||||| |||| |||| .| ||||
Db     16 AGAAGAAATTTCTATGCAGCAAAGCAATTCAAGAGTTGACTTGGGTGCTGTTAAAGGCATT 75

Qy    450 ywsntyathngnngargcngarcayaarwsnwsngayncayccngayaaaygtnat 509
          |.||.| || ||||||.|.|.| |||.||||.||||||.||| ||||||||.||||
Db     76 CAGATTTATAAGGGAAGCAGAGCATAAAAGTTCAGAAAATTCCAGCCTGACAATGTGAT 135

Qy    510 haaraaraaraaaycnttywsngargnaaarttyaaaytyngcngcngarathtgyathtg 569
          :|||.||:||||.||:||.||:||||||.|||||.|||.|.|.|.||:|||||||.|||.
Db    136 AGAGAAAAAACCCATTTTTGAGGAGAAAATTCAAGCACGTGCATAAAATTTGCATAAG 195

Qy    570 yaaygargaryghaaygtnaaytgncncargayaayayathwsntggacntgyca 629
          :|||||||||||.||||||.||| ||||||.|||.||.||| |.||||.|||:|.|
Db    196 TAACAAGAAGCCGCATGTTAATCCCAAGACAACAGACAATCTCCAGGACATATCA 255

Qy    630 rmgnwsnscarcarwsnathaarwsnytngcntgngnccnmgnmgnaatgggtytytg 689
          :.|:||:|||.|||.||.|.|.||.|||||.||.| .| |.||.||.|||.|||.||.
Db    256 GAGGTCTTCACAGCACCCTCCGTCACAGGGCCCTAGGAGAGAAATGGGTTTCC 315

Qy    690 yggnaacnggngccnggnwsnytnt-------gytcgygtncarccnmgngayytngtncc-- 740
          .|| |.|| ||.|| ||||.|.|:| |.||.|||||.||| .|| ||.|||:|.|||:||:
Db    316 TGGACCTGTGCAGCCTAGGTACTTGGCCACCCGTGTCCCAGTACTCCAGTAGTGGGCTGA 375

Qy    741 ---ntggytnccngtnaaytggncnwsngargngcnwsnccnaarccntggca 797
             |||||.|||.|||||.||||.||||.||.|.|||||.||.|||||||||
Db    376 AAGGGGCCAAAGTACACCTTACCCCATGGCTTCAGAGGGTGCAAGCGCCAAGCCTTGGCA 435

Qy    798 rytnccnwsngngtngarccngtngngcnaaraarwsnmgnathgargtntgggarcc 857
          |.||||.||.|.|||.|.|||.|.|||.|.||.||.||||.|||.||.|||.|||||
Db    436 GCTTCCCGCCATGCTGTTGAGCCTGCAAATGCACAGAATTTGGATTTGGGAACC 495

Qy    858 nccnathmgnttycaraaratathtayggnaayccntggatgccnmgncaraaarttygcngt 917
          .||.||.|| ||.|||.|||.||.||| ||.||.||||.||||.|.|.||.||||.||||.||
Db    496 TCTGCCTAGATTTCAGGAGAATGTACGGAAGTGTCCGGATGCCCAGGCAAAAGTTTGCTGC 555

Qy    918 nggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraaargnaaaygtnggntg 977
          .||.||.|| ||.|||||.|.||.||.||:.|||||.||.|||.|||||.|||||||.||
Db    556 AGGGGCAGGGCCTTCATGGAGAACCTCGCTGTGGAAGGGAAACATGGGGTC 615

Qy    978 ggarcncncnccnaaymgntnccnwsnsmgncngtngnwsncng---------------- 1037
          || ||.|.|.|||.| ||.||||.||..||.||.||.|||||.|| .||
Db    616 TGAGCCCCCACACAGAATCCCTACTGAGGCACCTGTGAGGGACTGTGAGGGC 675

Qy   1038 nccnwswsnmgnytncaraarggnngnwsnacngaywsnytncarcayytnccngaraa 1097
          .||.|| |||||.||.|||||||||| ||.||| ||.||||.||.|..|||||.|||||
Db    676 ACCATCTCCACACCCAGAAATGAACTGGAAGATCCACAACAGCTTGCACCATGCACCCTGGTG 735

Qy   1098 rwsnacngayacncartgycarccngtnaargncngcnggnatgg 1141
          :||.||.||||.|.||.||:||.||.|||||.|.|.|||.|||||
Db    736 AGCCGCAGACAATTCAACACCACGCCCGGGGGAAGCCGCAAAAGG 779
```

Right column:

```
DEFINITION  AGENCOURT_6478314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5563049
            5', mRNA sequence.
ACCESSION   BM471183
VERSION     BM471183.1  GI:18520225
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1036)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12293 row: g  column: 18
            High quality sequence stop: 661.

FEATURES             Location/Qualifiers
     source          1..1036
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5563049"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb. Library constructed by Life
                     Technologies."
BASE COUNT    254 a    239 c    294 g    247 t    2 others
ORIGIN

Query Match      23.4%;  Score 321.8;  DB 10;  Length 1036;
Best Local Similarity   45.0%;  Pred. No. 2.1e-62;
Matches 390; Conservative 166; Mismatches 267; Indels 43; Gaps 6;

Qy    553 gcngarathtgyathtgyaaygargaryt--naaygtnaayccncargayaaygngara 610
          ||.|.|.||.||.||||.||||||||.| ||.||.||.||||.||||.||.|.|.||
Db      1 GCAGAAAATTACATAAGTACGAAGGAACTACTAATGTTAATCCCCAAGACCATGGGGAAA 60

Qy    611 ayathwsntggacntgycarmgnwsnwsncarcarwsnathaaarwsnatgngc 670
          |.||| |||||.||||.|.||.||.||.||.|||.||.||||.|.||.||:.||
Db     61 ATGTCTCCAGTTCCTGTCAGGAGACCTTCACTGCAGAACCCTCCCATCACAGGCCCAGAGGC 120

Qy    671 cnmgnmg---naartggtytgyggnacnggngccnggnwsnytngtgtngtncarccnmm 727
          |.||.| ||.||| |:|.|| ||.|| ||.|| ||||.|.||||.|.|||.||.|.
Db    121 CCAGGAGAAAAGTGGTTCGTGTGGGGCTGCCCATGCTGTGTGGAAGCCTA 180

Qy    728 gngayytngtnccntggytnccncng---------------- 754
          |.||:.|.||||.|.|||.|.|.||| .||
Db    181 GGGACTTGGTGCCCCGTTGTCCCACGGTTCTCCAGCTGTGAAAGGGGTTAATGTATAG 240

Qy    755 aywsngcngtngcnwsngargggncgncwsnccnaarccntggcar---ytnccnwsngggng 811
          ||.||.|.||.|.| ||||||||| ||.|| ||.|||||||||||| |||||.||.|| |
Db    241 CTCTGGCTGTGGCTTTAGAGGGTGGAAGGGCTGGAAGCCCCAAGCCCCAACTCCTGGGTG 300

Qy    812 tngarccngtnggngcnaaraarwsnsmgnmgnathgargtntgggarccncncnccnaaym 871
          ||.|.||.||.||.|.|.|||||.|||.||||.|.||.|||.|||| |.|.|.|||.||
Db    301 TTGACCCTGTGGGGTGCCACAGAAGTCAAGAAACTGAGGTTTGGGAACCTCTGCT-AGATTCC 359

Qy    872 araarathtayggnaayccntggatgccnmgncaraaarttygcngtngngtnggntgggnwsnw 931
          |.||.|||.||| ||.||.|||||||||.|.||.||||.||.||.||.|.|.||.|| ||||
Db    360 AGAAGATGTATGGAAATGCCTGGATGCCCAGGCAAAAGTTTGCTCGTCAGGACCTCTGCT-AGATTCC 419

Qy    932 sntggmgnacnwsngcnmgngtngtncaraaargnaaaygtnggntgggarccncncncaym 991
          ||.|| |.||:.|| ||.||||.||.|||.|||||.|||||||.|| || ||.|.|.|.|
Db    420 AGAGAATGTATGGAAATGCCTGGATGCCCAGGCAAAAGTTTGCTCGTCAGGACCTCTGCT-AGATTCC 419
```

```
RESULT 7
BM457166
LOCUS       BM457166     898 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427
            5', mRNA sequence.
ACCESSION   BM457166
VERSION     BM457166.1  GI:18506206
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12346 row: h column: 20
            High quality sequence start: 9
            High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1..898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5583427"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      225 a    228 c    254 g    191 t
ORIGIN

Query Match   23.1%;  Score 318.6;  DB 10;  Length 898;
Best Local Similarity  45.1%;  Pred. No. 1e-61;
Matches 389;  Conservative 162;  Mismatches 253;  Indels 59;  Gaps 5;

Qy    562 tgyathtgyaayrgaryrgtnaaygtnaaycccnaargaryaaayathwsntgg 621
          ||:|||||||||:|  |:|:||||||||||   |:|||||||| |||:|||||
Db     35 TGCATAAGCAGCAAGGAGAGCCTAAATGTTAATCCCAAGACCATGGGGAAAAATGTCTCCAGG 94

Qy    622 acntgycaarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgmg--n 678
          |  |:|  |:|||| ||: : :|||||  : ::  ||||  |||:| |||||||   n
Db     95 TCAYGTCAGAGAGACTTTCAYGCAGCCCCATCACACAGGCGCCCAGGAGGAGAAA 154

Qy    679 aartggtttytgygngnacnggngnwsnytntgytgyytcarccnmngayytngtn 738
          || ||:|||  | :|: |  || ||| ||: :|:||:| :||   :|  |||:|
Db    155 AAGTGGGTTTTGTGGGCTGGGGCCCCAGGGTCTCGTGTGTCAGCCTAGGGACTTGGGTG 214

Qy    739 ccntgyygtnccng----                                 tnaaywsngcngtn 765
          ||||| :||||||                                      |||:|| | ||:|
Db    215 CCCTGTGTCCCAGCTGCTCCAGCCATGGCTGAAAGGAGCCAACATAGAGGTTGGGCTGTG 274

Qy    766 gcnwsngarggngncnwsnccnaarccntggcaryytnccnwsnggngtngarccngtnggn 825
          |||  || ||:| |||  ||| |||| ||| ||| :|||| ||||| |||| |||||:||
Db    275 GCTTCAGAGGGTAGAGGCCCCAAGCAYTGGCAGCTTCCATGTGAYGTTGAGCCTGTGGGT 334

Qy    826 gcnaaraarwsmgnathgargtntgggaarcccnccnnathmgnttycarararatthtayggn 885
          ||| |||||| ||| :|||:|::|||||  |||| ||   |||:|| |  ||  :||||||
Db    335 GCAGAGAYGTCAAGAATTGAGAYTTGGGAAACCCTGCCTAGAYTTCAGAYTATATATGGA 394

Qy    886 aaycnctngtgatgccnmgncaraaartttygcngtnggngtnggnwsnsntggmgnacnwsn 945
          || ||  |||||  |||||  || |||||||||||:|||||:||||   | ||| | || |
Db    395 AACACCTGGAYGTGCCAGGCAAAAAGTTTTGCTYCCAGACGGGCCCTCATGGGGAACCTCT 454

Qy    946 gcnmgngtngtncaraargnaaygtnggngarccnccncaymgngtnccnwsnggn 1005
          ||   ||||||| ||| :|| |||||||  ||||   |  |:|||||||  |||
Db    455 CCTAGGCACGTGTGGAGGGAAATGTGGG-CAGAGCCCCACACAGAGTCCCTGTTGGG 513

Qy   1006 gcnccnwswsnmgngcngtnmgmgnwsnccncncnws- 1044
          |||   || | ||| | |:|||:|| |: :: :| |- n
Db    514 GCACTGCCCTAGTGGAGCTGTGAGAAGAGGCCCACAGTCCCAGAACGGTAGACCCCAGAAT 573

Qy   1045 wsnmgnytncararggngmgnwsnacngayaywsnytncarcaygtnccngaraarwsnacn 1104
          ||  |  |:||| |:|  |||  |  || || || :|  |||||  ||| | |||  | || |
Db    574 GGTAGACCCCAGAAYGGTAGATCAACCACCAACTTGCCATCTTGTGCCTGGAAAAGCTACA 633

Qy   1105 gayacncaartgycarccngtnaaargcngcngggaarwsngtnccntayaaaracngtn 1164
          || |  |||:|||  || | |||| ||  ||   | |:||:|||  ||| || |  |||
Db    634 GACACTCAACATCAGCGCTGAGAAAGCAGGAGGCGTGTACTCTGTCAAAGCCACA 693

Qy   1165 gtncngarytnacnaaracngtnggnathtayytnytncaygtgcaaygayytngaygtn 1224
          |||  ||  |||  |||||||  | | ||:||||| :|  ||:|: ||||| ||| | |:|
Db    694 GGGGCGCGAGCTGCACAGACCATGGGGAACCTTAYCTCTTGCATCAGCATTGAATGTG 753

Qy   1225 mgncaygggngtnaarmgngaycaytyggngncytnmgnttygaytgycnacacnggntty 1284
          |||||||| ||||| :|||| ||||:|||| | |:| |:||:|||| |  | || || ||
Db    754 AGACAGGGAGTCAAAGAGAGATCATTATYGACAGAGTCTTAAAATTTGACTGGCCACTGGATTT 813

Qy   1285 mgnaacntayatgggnccngt-nccnytntgyttyggncarttytyytnccnttyggnacngc 1343
          |||| | :||||| |||||| |||  |:|:|:||||  |||:||| :|  ||:|||| || ||
Db    814 CAGACTTGCATGGACCCCTGYNAACCCCTYGTTTTTGGCAAYTTCTCTCCCACTGGAACAGC 873

Qy   1344 ngtnttyacncaartgyytntayy 1366
          |||:|||  | ||:|| :||||||
Db    874 TGTATTTACCCAATACCTGTACC 896
```

```
RESULT 8
AG086951
LOCUS       AG086951     717 bp    DNA    linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-085H08.F, genomic survey sequence.
```

```
ACCESSION   AG086951
VERSION     AG086951.1  GI:16638753
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee male
            BAC Library clone:PTB-085H08.F.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 717)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbee@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
PRIMERS
   Sequencing: -21M13
LIBRARY
   Vector        : pKS145
   R.Site 1      : SacI
   R.Site 2      : SacI.
FEATURES             Location/Qualifiers
     source          1..717
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-085H08.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT    175 a    189 c    195 g    158 t
ORIGIN

Query Match             22.7%;  Score 313;  DB 12;  Length 717;
  Best Local Similarity   49.1%;  Pred. No. 1.7e-60;
  Matches  299;  Conservative  113;  Mismatches  195;  Indels   2;  Gaps   1;

Qy    744 ygtnccngtnaaywsngcngtngcnwsnaargngngcnwsnccnaarccntggcarytncc 803
          ::|| | | |  ||| ::|| | || | ||: ||  | |  ||| |||||| |   || |
Db     89 cgtccgccgggagcrtcaggcgttcggccttcagaggtgcaaggcccttaagcctcttcc 148

Qy    804 nwsnggngtngrccncngtnggggcnaaraaarwsnmgnnathgrgtntggggarccnccnat 863
          ||:| | || |::| | |||:|| || ||||||: |: |  | :| | ||||| || ||||
Db    149 acatggtgttaagcctgtgagtccacagaaatcaagaacyagggtttggggaacctctgcc 208

Qy    864 hmgnttycarararthtayggnaayccntggatgccmngncaraarttygcngtngngt 923
          :|| ||| |||| | :|:||| || ||: ||| |||  |||||||:| | ||:|| ||:
Db    209 tagattycagaggatatatggtraataacctgggaatgcccagacacagaatytttgcytgtagggga 268

Qy    924 nggnwsnwsntggmgnacnwsngcnmgntngtncaraargngnaaygtngngtgggarcc 983
          ||| |: ||:|| :|| | |: || || | |||||| || || | ||:|| ||||| ||
Db    269 ggggggccttcatgaaaaacctcttcctagggcacgtgtggggagagaaaatgtggggccactatt 328

Qy    984 ncncaymgngtncccnvsnggngcncnnvsnggngcngtnmgngngmgnwsncncncnws 1043
          |:|:|  || ||:||  :|||| ||::  :||||| | | || || :| ||||:|:| |:
Db    329 atcacacagagtccctactggggcgcacccgctgtgaaaagtgggccactatt 388

Qy   1044 nwsnmgnytncararggnamgnwsnacnangayvsnytncarcaygtncngaraarwsnac 1103
          |:|: || || |||| | || |:|| || |:||  |:|| || |||| | |||||:|| ||
Db    389 ctccagaccccagaaatggtagacttccaacatccaccactggaaaagccac 448

Qy   1104 ngayacncaartgycarccngtnaargcngcngnggnatggarwsngtncncntayaaracngt 1163
          ||| ||| |||| |||| ||:|| ||:|| ||:| |||||| || | |:| || || ||:||
Db    694 TGGTCTGGACCAAGGTCCCCATGNGTTGTGTGCAGCCTAGGACTTCGCTGCTGGTGTCCCA 635
```

```
RESULT  9
BI754555/c
BI754555
LOCUS       BI754555                736 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603023384F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194028 5',
            mRNA sequence.
ACCESSION   BI754555
VERSION     BI754555.1  GI:15746133
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11485 row: g column: 21
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1..736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5194028"
                     /clone_lib="NIH_MGC_114"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
                     /lab_host="DH10B"
BASE COUNT    163 a    196 c    185 g    192 t
ORIGIN

Db    449 AGACACTCAATGCCAGCCTGTGAAAGCAGCAGCCTGGGAGGGAGGAGACTGTACCATGCAACGCAGACCAC 508
          || | |  |||||||| | ||:||  |||||||| |||||||| |||||||| || |||  ||||||| |
Qy   1164 ngtngcngarytnacnaaracngtnggnathtayytnytncaaytgycaygayyytngaygt 1223
Db    509 AGGGGCAGAGACCTGCCCAAAGACACCATGGGAACCACACTCTTGGCATACATCATGACCCTGGATGT 568

Qy   1224 nmgncayggngtnaarmgnaaycaytcytyggngngcnytnmgnttyagaytgycnacngngntt 1283
          ||| ||:||:|| ||:|| ||||| ||: |||| |||  || || :| |: |  |:| ||:|  :|
Db    569 GAGACATGGCACCAAAGGAGATCATTTTGAGCTTTTAAGATTTGACTTCCCTGCTGGAATG 628

Qy   1284 ymgnacntayatgggnccngt--nccnytntgttyggngncarttcyttycnttyggnacn 1341
          :|| ||:|| ||||| || ||  ||| :|:|||||| ||| ||||| || || ||| |||
Db    629 TCAGACTTGCATGGGCCCTGTAGCCCTGCTTTTTTTTTGGCCAAATTCCCCATTTGGAATG 688

Qy   1342 gcngtntty 1350
          || ::| |:
Db    689 GGGTAATT 697
```

```
RESULT 10
LOCUS       AG112666       654 bp   DNA    linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-119B01.R, genomic survey sequence.
ACCESSION   AG112666
VERSION     AG112666.1  GI:16733185
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC library clone:PTB-119B01.R.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       BAC end sequences of Library PTB
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 654)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe.H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045,Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.

FEATURES             Location/Qualifiers
     source          1..654
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-119B01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      165 a    174 c    180 g    135 t
ORIGIN
```

Query Match 21.8%;   Score 300.4; DB 12;   Length 654;
Best Local Similarity 49.3%; Pred. No. 1.2e-57;
Matches 296; Conservative 112; Mismatches 191; Indels 2; Gaps 2;

```
Qy    719 tncarcnmgngayytngtnccntgygttccngtcnaaywsngcngtngcnwsngaarggng 778
          | |:| |||||  | |:|
Db    634 GCCACTCCAGCCATTGCTAAAAGGGCGCCAAGGTACAGCTTCAGCCCATGGTTCAGAGACT 575

Qy    779 cnwsncnaarcncntgycarytnccnwsnggngtngarccngtnggngcnaaraarwsnm 838
          |:|  |||||||||:| |:||| ||| |||||||||
Db    574 GCAAACCCCAACCTTGGCAGCTACCCATGTGGTCGGGTGTACAGAAGTCAA 515

Qy    839 gnathgargtntgggarccncncnaathmgntycaraararathtayygnaayccntggatgc 898
          | :|:||||||||||||| |||||||||||  :|:|||||||
Db    514 GAATTGAGGTTGGGAACCTCCATCTAGATTCAGAAGATGTRATGAAAATGCCTGGATGC 455

Qy    899 cnmgncaraarttygcngtnggggntnggngtnacnwsngcnmgntngtngtnc 958
          ||| ||||| |: | |||||||||| ||||  |:||||  ||
Db    454 CCAGGCAAAAGTTTGCTCCAGGGGCGAGGGGCCCCTTGGAGAACCTCTCGCCAGTGC 395

Qy    959 araargnaaygtnggntgggarccncncncaymgngtncccnwsnggngngcnccnwsnwsnm 1018
          ||||||||||||  ||||||||||||||||||||||||||
Db    394 AGAAGGGAAAATGTGGGGTCAGAGCCCCAACACTAYATCCCTACYGGGGTACTGCCTAGTA 335

Qy    1019 gngcngtnmgngmgnwsncccncncnwsnwsnmgnytncaraaarggnmgnwsnacngaywsny 1078
          ||: ||||||| ||||||||||||| |:|||||||||||||
Db    334 GAGCTGTRAAGAAGAGCCCACCACATCCTCCAAACATCAGAATGGTAGATACACTGACAGCCT 275

Qy    1079 tncarcayytnccngaraarwsncngayacncarttgycaarccngtnaargcngcngcngna 1138
          ||||||||:|||||| :|||| |||||  ||||||||  |
Db    274 TGTACAAGGTGCCTGGAAAACGCGCAGACACTCAATGCCAGCCCATGAAAGAAGCCAGCA 215

Qy    1139 tggarwsngtncntayaaaracngtngtcngaryytnacnaaracngtngnathtayy 1198
          ||||||||||||:|  | | ||||||  || : ||  | |:||||
Db    214 GGGGAGCTATGCCCTGCAAAAGCCACACAGGGGTGGAGCTGCCTACCGACTATGGGACACAC 155

Qy    1199 tnytncaytgycaygayytngaygtnmgncaygggngtnaaramgngaycayttygggngcny 1258
          |  ||||||  ||| |:|:|| ||||||||  |:|||:|| |||
Db    154 TCTTGCATCACTCTGACTTGGATAATGAGACAATCCAGTCAAAAGAAATCACTTTGGAGCTT 95

Qy    1259 tnmgnttygaytgyccnacnacngnttymgnacntayatggnncnacngtnccnytntgyttyg 1318
          ||||||||   |:|||:| |||||||  |:|
Db    94 CAAAAATTTGACTGCCCCGCCTGGATTTGCATGGGGTCCTCGTGTAACCCCTTTGTTTT 35

Qy    1319 gncaarttyyttycyccnttgggnacngcngtnttyac 1352
          |||||||||||| :||||:||  |||||| |||||
Db    34 GGGCATTGTCTCCCATTTGGAATGGCCGTATTCAC 1
```

```
Qy    760 gcngtngcnwsngarggngcnwsnccnaarccngtngcaryytncnnwsngnntggntngacg 819
          | |||||:|:||||:|||||||||||||||   ||| |||||:|||| :|:|||
Db     50 GCTGTRAGCTTCAGAGGGTGCAAGCCTCAAGCCTGGCAGCTTCCACATGAATGTTAAGCCT 109

Qy    820 gtnggngcnaaraarwsnmgnaathtgargtntgggargttcnyggngayytnggayytnga 879
          ||||||||| ||:| ||| ||| :|||||| ||| |||||||  | | |:|||:||
Db    110 GTGAATGCACAGAAGTCAAGAAGTTTGGGAACCTCACCTAGAATTTCAGAAAATG 169

Qy    880 tayggnaayccntgygatgcnmgncaraartttcgngtngggtnggngwsnwsntggmgn 939
          ||||||||:||||| ||||||||  ||:|:||||||||||||||
Db    170 TATGGAAACTCCTGGATGCCCAGGCAGAGGAGTTTGCTCCTGGGGTGGGGCCCATGGAGA 229

Qy    940 acnwsngcnmgngtngtngcngttncaraargnaaygtnggntgggarccncncnacaymgngtncc 999
          |:||||||||  ||||:|||||||:|||||||||||
Db    230 AACTCTGCTAGGGCAGTGCAGAAGAAAACGTGGGGGTGGGAGCGCCCATCACAGATGTCCCT 289

Qy    1000 wsngcngcncnwsnwsnmgncngtnmgnwsncncnccnwsnwsnmgnytncaraar 1059
          |:||||||||:||||||||| :|| : |:|| |:|||
Db    290 ACTGGGGCCACTGCCTAGG-GAGCCGTGAGAAGAAGAGGCCCCATCAGTCCTCCAGACCCCAGAAT 348

Qy    1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayancarttgycar 1119
          |||:||||| : |||||| || ||| |||||| :|:| |:|:|||
Db    349 CGTAGACCACTGACAGCTTCGCCCACATGTACCTGGAAAGCCAACACTCAACACCAC 408

Qy    1120 ccngtnaargcngcngnnatggaarwsngtncctayaaaacngtngtgngcngaryytnacn 1179
          || |||||| |:| |:| ||||| |:|:| |||||:|||| |:|||:||  |:|||
Db    409 CCCATGAAAGCCAACTGGGAGAGAGCCTGTACCCTGCAAAAGCCACTGAAGTGTGGCC 468

Qy    1180 aaracngtnggnaathtayytnytncaytgycaygayytngaygtnmgncaygggngtnaar 1239
          ||:| ||||||| ||:||:|||| |||||| |:|||||:|||
Db    469 AAGACTAGGGGAACCTACCCTTGCATCAGCATGACCATGAGCTGGATGTGAGCACGGAATCAAG 528

Qy    1240 mgngaycaytgygggncycntgnmgnytgaytgyccnacngngnttygnacntayatgggn 1299
          ||| |||||| |: |::|||||:| ||| || |: | |||||| ||:||
Db    529 GGAGATCACTTTGGAGCTTTAAGATTTGACTGCCCCGCCTGGATTTGCATGGGC 588

Qy    1300 ccngt-nccnytntgyttyggncaattyttyccnttyggnacngcngtnttyacncartg 1358
          |||||  :||||:| |||||:||||||||:|:||||||||||||||| :|||| ||||
Db    589 CCTGTAGCCCCTTAGTTATGGCCACATACCTCCCATTTGGAATGGCTGCACTTCACCCATG 648

Qy    1359 y 1359
          :
Db    649 T 649
```

```
RESULT 11
LOCUS       AG064424       676 bp   DNA    linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-053H24.R, genomic survey sequence.
```

Left column:

```
ACCESSION   AG064424
VERSION     AG064424.1  GI:16616226
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-053H24.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 676)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
FEATURES             Location/Qualifiers
     source          1..676
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-053H24.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
     PRIMERS
                     Sequencing: M13Rev
     LIBRARY
                     Vector     : pKS145
                     R.Site 1   : SacI
                     R.Site 2   : SacI
BASE COUNT       164 a      165 c      191 g      156 t
ORIGIN
```

```
Query Match                        21.5%;   Score 296;  DB 12;   Length 676;
Best Local Similarity   47.9%;   Pred. No. 1.3e-56;
Matches  294; Conservative  117; Mismatches  201; Indels    2;  Gaps   2;

Qy   760 gcngtngncwsngargnggngcnwsnccnaarccntggcaarytnccwsnggngtngarccn 819
         |||||| |:|| | |:|| |:: | :|:|||| | | :|| |:| || ||  ||
Db    63 GCCATGGCGTTCTAAAGGTGCAAGCCTCAAGCGTTGGCACATGATGTTGAGCTG 122

Qy   820 gtnggngcnaaararwsnmngnnatngargtgntggargarcncncnathmgnttycaraaarath 879
         | | |||||:: ||::|| | |  ||| || ||  | | || |  ||:| :||
Db   123 GCAAGTGCACAGAAGTCAAGAAACTGGGGTTTAGGAACCTCTACCTTAGATTTCAGGAATG 182

Qy   880 tayggnaayccntggatgccnmgncaaraarttygcngtngngtngggnwsnsntgggmgn 939
         || :||  :|  :|::|  :| || :|:|::|||| :| |  ||:| |:|| |||:|
Db   183 TATGGAAATGCCTGGATGTCCAGCGACAGAGTTTCCCGCAGGGAGTCCCTGGACGGGAGA 242

Qy   940 acnwsngcnmgngtngtncaraarggnaaygtngngtnggggarccncncncaymgngtnccn 999
         || :|:| | |::| ||:  :|| || |  |||:|||:|  ||:|:| |  || |
Db   243 ACTCTGCTAGGGACAAATGTGGAAGGGAAATGTGGGGTTGGAGTCCACGCCCAGAGTCCCT 302

Qy  1000 wsnggngcnccnwsnwsnsmngcngngtnmgnmgnwsnccncncnwswsnsmngnytncaraar 1059
         ::|::|| |  |  || |::: | |:|| : ||  |: |:||| | :|||||
Db   303 ACTGGGGCACCATCTAGTGGAAG-AGGGCACTGACCTCCAGACCCAGAAT 361

Qy  1060 ggnmgnwsnacngaywsnytncarcaaygtnccngaraarwsnacngayacncartgycar 1119
         || :: ::|:  || ||| :| ||: |:||  :|  | | :|| | | | |:
Db   362 GGTAGAATCCACTGACAGCTTACACCATGTGCCTCGAAAAAGCTGTAAACACTCACACCAG 421

Qy  1120 ccngtnaargcngcngngnatggarwsngtnccntayaaracngtngtngcngnarytnacn 1179
         || |:|| | || | | || |  || | | :|::| | | | | ||  | | |
Db       |
```

Right column:

```
Db   422 CCCATGAAAGCATCTGGGAGGGAGGGAGGTTGTTACCCTGCAGATCCACAGGGGCGGGGAGCTGCCC 481

Qy  1180 aaracngtnggnathtayytnytncaytgtgycaygaytngaaygtngmgncaygggngtnaar 1239
         |:||| :|||  |   |   :||:|:| |:|  ||:||:|  ||||| |:| :|||
Db   482 AAGACCATGGAGGCCCAACCTCTTGCATCAGTGTGACCTGGATGTGAGACATGGAGTCAAA 541

Qy  1240 mgngaycaytygngncaytnmgnttygayytgaytgyccnacngngntttygnaactayatgggn 1299
         | |:|: ||:|| |:|| :|||:|| | |||:||| | | |::| | :|| |||| :||
Db   542 GGAGATCATTTGTAGTTTAAGATTTGAACTGCCTCCCTGGATTTGGACTACGACATGGG 601

Qy  1300 ccngt-ncnytntgtttyggncaartytyttyccntyggnaacngcngtnttyacncaarty 1358
         ||:|| |:||| ||||| :|||| |:|| |||:| :||| || | | :| ||:|
Db   602 CCTGTAGCCCCTTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGCGTCGTTATTTACCCAATG 661

Qy  1359 yytntayytncayt 1372
         ::|| ::| ||:|
Db   662 CCTGTACCCCATT 675
```

```
RESULT   12
LOCUS       AG062447             691 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-050M20.F, genomic survey sequence.
ACCESSION   AG062447
VERSION     AG062447.1  GI:16614249
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 691)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
FEATURES             Location/Qualifiers
     source          1..691
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-050M20.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
     PRIMERS
                     Sequencing: -21M13
     LIBRARY
                     Vector     : pKS145
                     R.Site 1   : SacI
                     R.Site 2   : SacI
BASE COUNT       160 a      167 c      201 g      162 t      1 others
ORIGIN
```

```
Query Match                        20.9%;   Score 287.4;  DB 12;   Length 691;
Best Local Similarity   49.0%;   Pred. No. 1.2e-54;
Matches  301; Conservative  110; Mismatches  196; Indels    7;  Gaps   3;

Qy   744 ygtnccngtnaaywsngcngtngcnwsngnargygngcnwsnccnaarccntggcaarytncc 803
         :|||   || ::: :| :||| :||| | |||| | ||:|::|| ||| |  | |||
Db    82 CGTCCGCGGGAGCTCAGGCTGTGGCTTCAGAGGGTGCAAGACCCAAGCTTTGGAGGCTTCC 141
```

RESULT 13

LOCUS   AG116938   720 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-124K01.R, genomic survey sequence.
ACCESSION   AG116938
VERSION   AG116938.1 GI:16737457
KEYWORDS   GSS; GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-124K01.R.
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   BAC end sequences of Library PTB
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 720)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: M13Rev

FEATURES   Location/Qualifiers
     source   1..720
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-124K01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
LIBRARY
     Vector   : pKS145
     R.Site 1   : SacI
     R.Site 2   : SacI
BASE COUNT   181 a   186 c   187 g   165 t   1 others
ORIGIN

Query Match   20.8%; Score 286.2; DB 12; Length 720;
Best Local Similarity   48.5%; Pred. No. 2.3e-54;
Matches 298; Conservative 112; Mismatches 200; Indels   4; Gaps 3;

Qy   804 nwsnggngtngarccngtnggngcnaaraarwsnnngnathgargtntggarcncncnat 863
          ::|| ||::||| || ||| || |:|:|:|| |:||::||||:|| | | ||| |
Db   142 AtGTGtGTGTTgGAGCCTGTGAGTGCACAGAGA-----AtCGGGTGTTtGGAACCTCCACT 196

Qy   864 hmgnttycaraaraththayggnaaycncntggatgccngmncaraaarttygcngtnggngt 923
          ::|| ||||||||||||||| || ||:|:|| || || ||||||||||||||||| ::||
Db   197 TAGAtTTCAGAAGAtGTAtGAAAAtACCTGGAtGCCATGCAAAAGTTTGCTGTAGGGTC 256

Qy   924 ngguwswsntggmnacnwsngcmngngtngtncaraarggnaaygtngngntgggarcc 983
          |||| || ||||||||||||| ||||||| || ::||||||:||:||:|| |||||||
Db   257 AGGGGCCTCATGGAGAGAACTTCTGCTAGGGCAGTGTGGCAGGGAAAtGTGAGGTCAGAGGC 316

Qy   984 nccncaymgngtnccncnwsngngcncccnwsnwsnmgncngtnmgmgnwsnccncnws 1043
          || || ||:| |:||| || ::|:||||:||| ||||| ::||||| |: |||
Db   317 CCCACACAGAGTCCCTACTCGGGCACCACCTGTGAGCGTGAGAAGAGGCCACTGTC 376

Qy   1044 nwsnmgnytncaraargngmngnwsnacngaywswsnytncarcaygtnccngararwsnac 1103
          ||:| ||| |||||||:|| | |:|||:|||:| | ||| ||:||||: ||||||||:||
Db   377 CTCCAGACTCCAGAAAGGTAGAAtCCATCGATGGCTTGCATCATGGTCCCTGAAAACCCAC 436

Qy   1104 ngayacnaartgycarccngtnaargncngngnatggarwsngtnccntaayaaacngt 1163
          |:||:| |||:|:|| ||||||||| |||||||||| |||||||| || || |||||||
Db   437 AGACACTCAAtGCCAGCCCAtGAAAGCACGCCATGAAAGCAGCCTGTACCCTNACAAGGCCAC 496

Qy   1164 ngtngcngaryt-nacnaaracngtnggnnathtayytnytncaytgycayggayytngayg 1222
          || |||| ||:| || | ||||:|| || || |||||||| || :||||||:|| |||||:
Db   497 AGGGGCGAGCGTTCCAAGACCATGGGAACCCACCCTCAGTGTGACCTGGTTG 556

Qy   1223 tnmgncaygngtnaarmngaycayttyggngcnytnmgnttygaytgycnacacnggnt 1282
          | || |||:|||| |:||||||||||||| ||||||||||| ||:|||:||| |||
Db   557 TGAGAtATGGAGtCAGAGGAGAtCATTGTAGAGCTTTAAGAtTTGACTGTGTGGCTGGAT 616

Qy   1283 tymgnacntayatgggnccngt-nccnytntgyttyggncaarttyyttycnttyggnacn 1341
          ||||||||||||||||||||||| |||||||:||| ||:|||:||||:|:|||||||||||
Db   617 TTGGGACTtCCATGGGGCCGTGAGTCCTTGTTTGGGCAATTTCTCCAATTGTAAtG 676

Qy   1342 gcngtntttyacnca 1355
          |||| ||:|||| ||
Db   677 GCTGTATTtACCA 690

RESULT 14

LOCUS   AQ037711   623 bp   DNA   linear   GSS 11-JUL-1998
DEFINITION CIT-HSP-2337G10.TR CIT-HSP Homo sapiens genomic clone 2337G10, DNA
            sequence.
ACCESSION   AQ037711
VERSION   AQ037711.1 GI:3303543
KEYWORDS   GSS.

Qy   760 gcngtngcnwsngargngngcnwsnccnaarccntggcaarytnccnwsnggnggtngarccn 819
          ||||| ||||||||:|:||||||| ||:|||| ||||:||||| ||||| || ||||||||:
Db   110 GCCATtGGCTTCACAtGGTGTTCCAACCACCACAAGTCTTtGGTGTGGCTTCCACATGGCATTCCA 169

Qy   820 gtnggngcnaaaraarwsnmgnathtayytnytncaytgycaygayytngayg 879
          || |||||||||||||||||||||:|||||||:|| |||||:||:| | |:|
Db   170 GCGGGGTGCCACAGAAGTCAAGAAtTGAGGTTtGGAAACCTCCACCTGAGTTCAGAACATG 229

Qy   880 tayggnaayccntggatgccnacnggnttygaytgycnacnggnt 939
          ||:||||| | ||||| || ||| || |:|| |||||| | || |
Db   230 TATGGAAAAtGCCTGGAtGCCCACAGCCAGCAGAGTTTGCTGCAGGGAGTGGCTCTCAGAGA 289

Qy   940 acnwsngcnmgngtngtncaraargngnaaygtnggnttgggntnmgntgnggcngtncn 999
          || || || ||||| || |:|||||:|:|||||||: |:|||:|:|||||| ||||:||
Db   290 ACCTCTGCTAGGGCAGTGCAGAAtGGAAGGGAAAtGTGGGGTCGGAGCCCCCACACAGGGTCCT 349

Qy   1000 wsnggngcnccnwsnsnmgncngtnmgmgnwsnccncnwsnwsnmgnytncaraar 1059
          || ||||||||||| || ||| ||| ||| || | || || ||||:|| || |:|
Db   350 ACTGCGG-AtCACCTAGTGGAGCTGTGAGAAGAGCAGCCACTGTCTTCCAGACATCAG 408

Qy   1060 ggnmgnwsnacngaywswsnytncarcaygtnccngararwsnacngaycacnartgycar 1119
          ||:||| |:| | :| | ||| | |||:: || ||||||||:| ||| | || |:|:
Db   409 GCTAGAtCCACTCAACACTTGC--ACCTGTGCTTCGGAAAAGCTCCAGGTACTCAACAtCAG 466

Qy   1120 ccngtnaargncngcngnatggarwsngtnccntaayaaaracngtngtngcngararytnacn 1179
          || |||||| |:|||||||||| |||||||| ||:||| ||:|||||| |:| | |||||||||
Db   467 CCCATGAAAGCACGCCATGAAAGCAGCCTGTACCCTGCAAAGCAACAGAGGTGAGCTTCCC 526

Qy   1180 aaracngtnggnnathtayytnytncaytgycaygayytngayggngcnaaaraarwsnmgn 1239
          || | ||||| ||| |||||||| || :||||||:||:| ||||| |||||||||||||||
Db   527 GAGACCAtGGGAAACCTACCTCTTCCATCAGGGTGACCTGGATGTGAGACAtGGATTCAAA 586

Qy   1240 mgngaycayttyggngcnytnmgnttygaytgycnacnggnttyymgncaacntayatgggn 1299
          ||:|||||||||||||| |||||||||| ||:|||:|||||||| |||||||||||||||
Db   587 AGAGATCAtAtTGGGAGCTTTAAGAtTTGACAGCCACACTGGAtTTTtGGACTtCCATGGGC 646

Qy   1300 ccngt-nccnytntgyttyggncaarttyyttycnttyggnacngtntg 1358
          |||||| |||||||:||| ||:|||:||||:|:|||||||||||||||||
Db   647 CCTGTAGCCCTtTGTTTGGGCAATTTCTCCAATTGTAAtGGTTGCAttTACCTAAtG 706

Qy   1359 yyntatayytncayt 1372
          ::|| ||::|| |:|
Db   707 TCTGTAtTCCCAtT 720

```
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 623)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
TITLE       Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
JOURNAL     Unpublished (1998)
COMMENT     Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Contact: Mark Adams
            The Institute for Genomic Research
            Department of Eukaryotic Genomics
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Other_GSSs: CIT-HSP-2337G10.TF
FEATURES            Location/Qualifiers
     source         1..623
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="2337G10"
                    /clone_lib="CIT-HSP"
                    /sex="Male"
                    /cell_type="Sperm"
                    /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                    HindIII"
                    Class: BAC ends.
BASE COUNT      160 a    146 c    172 g    145 t
ORIGIN

Query Match          20.6%;    Score 283.8;  DB 12;   Length 623;
Best Local Similarity   48.1%;   Pred. No. 7.3e-54;
Matches 286; Conservative 111; Mismatches 196; Indels  2; Gaps  2;

Qy   784 ccnaarccntggcarytncnchwsnggngtngarcccnngtnggngcnaataarwsnmgnath 843
         ::|:|||:|||||:|||||||||||| |||||| |||:|||||:||||||| |||:|::||
Db    26 CTCAAGCTTGGCAACTTCACATGGTGTTGAGTCGCGGTACACCAAAGTCAAGAACT 85

Qy   844 gargtntg-ggarccncnccnaathmgnttycaraarathtayggnaayccntggatgccnmg 902
         ||:|||| ||||||:|||||| || :||||||:|::|:||||||||||:||||||||:||||
Db    86 GAGGTTTGAGGAACCTCCACCTACATTTCAGAGACATGTATGGAAAATGCCTGGATGTCCAG 145

Qy   903 ncaraarttygcngtcngtnggngtnggngnacnwsnggncnmgntgntncaraa 962
         :|:||||||||||:|:||:||||:||||||: |||:||||:|||||:||||||||
Db   146 GCAAAAGTTTGCTTGCCAGGGCAGGGCCCAGGATGGAAGAACCTCTCCTAGGGCAGTGCGGAA 205

Qy   963 rggnaaygtnggngntgggggarccncncacaymgncngcnccnwsnwsnmgngc 1022
         :|||||:|||||:|:||||||||||:|:||| :||:||||:|||||:||||:|||
Db   206 GGGAAATGTGGGGTTGGAGCCGTCACAGAGAGTCTACTGGGGCACTGGTGGAGC 265

Qy  1023 ngtnmgnmgnwsnccnccnwsnwsnmgnytncaaraargggnmgnwsnacngaywsnyntnca 1082
         :|||:||:||||||:|:||||:||:||:||||||||||||| :||||||:||:||||:||:||
Db   266 TGTGAGAGAGGGGCCATTGTCCTCCAGCACCTAGAANTGTAGARCTGCCGCACACHTTGCA 325

Qy  1083 rcaygtnccngarcarwsnacngaycacncartgycarccngtnaargncngcnggnnatgga 1142
         :|||||:|||||:||:|||||:|||||||:||:||:|||||||||||:||:|||||:||||||
Db   326 TTATGCAACTGGAAAAGGCACHACACACCCCATGAAAACAGGCCAGGAGGGG 385

Qy  1143 rwsngtnccntayaaracngtngtcngaryntnacnaaracngtnggnnathtatayytnyt 1202
         :||:||:|||||||||:||:||:|:||:|:||||:||||:||:|||||:|||:|||:||:|:
Db   386 AGCCATNACCCTGCAAAGCCACACAGGTGCCCTGAGGCCCATGGGAACCCACCTCTT 445

Qy  1203 ncaytgycaygayytngaygtnnggcayttyggngcnaytnmg 1262
         :||||||:||||:|||||:|| :|||||||:|||||:|||:||
Db   446 GCATCAGCCTGACCCGAGATGTGACACATGGTGTCAAAAGGAGAGATCATTTTTGGAGCTTTAAG 505
```

```
Qy  1263 nttygaytgycnacnggngtntymgnacntayatgggnccngt-ncnytntgytgtyggnc 1321
         :||||:||||:|:||:||||||||:||:|||||||||:||||| :|:||||||||||||||
Db   506 ATTTGACTGCGCCTGCTGCGGATTTCAGACTTGTRAAGGGGCCCAGTAGCGCCCTTTTCTTTTGGCC 565

Qy  1322 arttyttyccnttyggnacngcngtnttyacncartgyyntnatayytncaytgyat 1376
         |:||:||:|||||:|||||:||:||:|||||||:|||:::|:||||||:|||||:||
Db   566 AATTTCTCCCATTTGGAATGGCTGTRATTYACCAATGCCYGTRATTCCCATTGTAT 620
```

```
RESULT 15
AG097258

LOCUS       AG097258     677 bp    DNA    linear    GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-098001.R, genomic survey sequence.
ACCESSION   AG097258
VERSION     AG097258.1  GI:16717775
KEYWORDS    GSS.
SOURCE      Pan troglodytes (genome survey sequence).
            Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-098001.R.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       BAC end sequences of Library PTB
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 677)
AUTHORS     Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: M13Rev
            LIBRARY
            Vector         : pKS145
            R.Site 1       : SacI.
            R.Site 2       : SacI.
FEATURES            Location/Qualifiers
     source         1..677
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-098001.R"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      178 a    175 c    176 g    147 t    1 others
ORIGIN

Query Match          20.6%;    Score 283.2;  DB 12;   Length 677;
Best Local Similarity   48.8%;   Pred. No. 1e-53;
Matches 294; Conservative 113; Mismatches 187; Indels  9; Gaps  3;

Qy   760 gcngtngcnwsngargggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 819
         ||:||:||:||:||:|||||||:|||||||:|||||||||||:|||:|||||:||:||:|:
Db    68 GCTGTGGCTTCAGAAGGTAGAAGCCCCAAGCGTTCCATGTGGTTGTTGAGCCT 127

Qy   820 gtnggngcnaaraarwsnmgnnathgargtntgggarccncnccnaathmgnttycaraarath 879
         ||:|||||:|||||:||:||: |||||:||||||||||:|:||||| ||:||||||:|:||||
Db   128 ACAGGTGCCACAGAGTCACAAATTGAGGTTTGTGAACCTCCACCTAGATTT-AGACGATG 186

Qy   880 tayggnaaayccntggatgccnmgncaraattygcngtnggngtnggnnwsnsntggmgn 939
         ||||||||:|||||||||||:||:|:||||||||||:||||:||||: ||||:||||:|:
Db   187 TATGGAAACACCTGGATGCCCCAGGTAAAAGTTTGCTGCAGGGTGGAGGCCATCATCGGAGA 246
```

```
Qy    940  acnwsngcnmgngtngtncaraargngnaaygtnggntggaarccnccncaymgngtnccn   999
           || :||  | || |  ::|   :| |  ||:|| |||| | |::|| | ||
Db    247  ACCTTCTTCTAGGGCAGTACGGAAAAGGAAATGCAGGGTTGGAGCCCCCAAACAGAGTCCCT   306

Qy   1000  wsnggngcnccnwsnwsnmngngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar   1059
           ::|| | || | |  : ||| |  : :|  : | || ||  |  ||||
Db    307  AATGGCGCACTGCCTAGTGAAGCTGTAAGAAGAGGACCACCATCCTCCAGACCCCCAGAAT   366

Qy   1060  ggnmgnwsnacngaywsnytncarcaygtnccngararaarwsnacngayacncartgycar   1119
           || |    :: | :|  || |||  :  :  |  ||   |::||:::||:
Db    367  GGTAGATCCACCTACAGCTTGCAACATGCACCCTGGAAAAGCTGCAGACACTCAGCGCCAG   426

Qy   1120  ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn   1179
           |:|| | || | |||||:|| ||::: | ||    |::|| |: |
Db    427  CCCATGAAAGCACGCCAGGAGGGAGACTGTTCTCCTGCAAATCCACAGGGGCCAGAGCTGCCC   486

Qy   1180  aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar   1239
           ||:||| | :| | | |:::| | ||||    :: | ||::||  :| |  ||:|
Db    487  AAGACCATGGGAACCCACCTCTTCGCATCAACGTGACCTGGATGTGAGGCATGGAGTCAAA   546

Qy   1240  mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttcymgnacntayatgggn   1299
           :| | |||:||:| ||:||:|:||||  | :|  |  | : | ::: | |   |  ||:|:
Db    547  GGAGATCATTNTGGAGC-------TTTGATTGCCCTGCTGGATTTTGACTTGCANTGGGC   599

Qy   1300  ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg   1358
           |||:|| :||:| |  : || | ||| | ||:::|| ||:|  ::| |:|| :|| |||
Db    600  CCTGTATCCCCTGTATTTTGAGCAATTTATCCCATTTGAAATGGCTGTATTTTACCCAATG   659

Qy   1359  yyt   1361
           ::|
Db    660  CCT   662
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 18:08:58 ; Search time 5686.68 Seconds
(without alignments)
6369.944 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg..........gyytntayytncaytcgyatg 1731
Scoring table: IDENTITY_NUC
  Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 1046268293 residues

Total number of hits satisfying chosen parameters: 3595312

Post-processing: Minimum Match 0%
  Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 200000000000

Database :   GenEmbl:*
Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Description

(Database columns listing)

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

No.	Score					Description
1	921	53.2	145880	9	HS302D9	Z82198 Human DNA s
2	497.8	28.8	154090	9	AC025577	AC025577 Homo sapi
3	497.8	28.8	186660	2	AC026107	AC026107 Homo sapi
4	493.6	28.5	124518	2	AL138962	AL138962 Human DNA
5	493	28.5	123070	2	AC087999	AC087999 Homo sapi
6	493	28.5	189768	2	AC044889	AC044889 Homo sapi
7	490.4	28.3	198545	2	AC017063	AC017063 Homo sapi
8	488.8	28.2	161575	2	AC025233	AC025233 Homo sapi
9	488.8	28.2	163674	2	AC087500	AC087500 Homo sapi
10	488.8	28.2	175639	2	AC015727	AC015727 Homo sapi
11	488.6	28.2	181842	9	AL391823	AL391823 Human DNA
12	488.2	28.2	77405	2	AL353634	AL353634 Homo sapi
13	488.2	28.2	125295	2	AL672061	AL672061 Human DNA
14	487.6	28.2	124531	9	HSJ633H17	HSJ633H17 Human DNA s
15	487.6	28.2	407114	9	Z69722	Z69722 Human DNA s
16	487	28.1	159712	9	AP004219	AP004219 Homo sapi
17	487	28.1	146743	9	AL359232	AL359232 Human chr
18	487	28.1	152544	9	AC093588	AC093588 Homo sapi
19	487	28.1	166679	9	AC079899	AC079899 Homo sapi
20	487	28.1	174662	9	AC026036	AC026036 Homo sapi
21	487	28.1	176426	9	AC007370	AC007370 Homo sapi
22	487	28.1	203726	9	AC011882	AC011882 Homo sapi
23	486	28.1	139005	2	AF286112	AF286112 Homo sapi
24	486	28.1	156458	2	AC025647	AC025647 Homo sapi
25	486	28.1	165901	2	AP003474	AP003474 Homo sapi
26	486	28.1	171427	2	AC021590	AC021590 Homo sapi
27	485.8	28.1	99084	2	AC026420	AC026420 Homo sapi
28	485.8	28.1	133790	2	AC012080	AC012080 Homo sapi
29	485.8	28.1	150332	9	AC004921	AC004921 Homo sapi
30	485.8	28.1	174293	9	AC008816	AC008816 Homo sapi
31	485.8	28.1	203043	9	AC017093	AC017093 Homo sapi
32	485.4	28.0	131215	9	AC079614	AC079614 Homo sapi
33	485	28.0	107885	9	AC006389	AC006389 Homo sapi
34	484.6	28.0	173480	9	CN500M8T	CNS00M8T Human chr
35	483	27.9	164370	9	AC097460	AC097460 Homo sapi
36	482.6	27.9	112515	9	AL139216	AL139216 Human DNA
37	482.6	27.9	142224	2	AC095535	AC095535 Homo sapi
38	482.6	27.9	178535	9	AL354978	AL354978 Homo sapi
39	482	27.8	168502	9	AC091005	AC091005 Homo sapi
40	481.6	27.8	38235	9	AC004559	AC004559 Homo sapi
41	481.4	27.8	234622	2	AC110089	AC110089 Homo sapi
42	479.8	27.7	187883	9	AC008268	AC008268 Homo sapi
43	479.6	27.7	167891	9	AC013439	AC013439 Homo sapi
44	479.2	27.7	73390	9	AC012038	AC012038 Homo sapi
45	479.2	27.7	94730	9	AP000230	AP000230 Homo sapi

(Note: entries 2-13, 37, 41 marked with "c")

ALIGNMENTS

RESULT 1
HS302D9
LOCUS   HS302D9   145880 bp   DNA   linear   PRI 12-DEC-1999
DEFINITION   Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains
GSSs, complete sequence.
ACCESSION   Z82198
VERSION   Z82198.2   GI:6572207
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 145880)
AUTHORS   Bridgeman,A.
TITLE   Direct Submission
JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
COMMENT   On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
RP1-302D9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: PCYPAC2
This sequence is the entire insert of clone RP1-302D9 The true left
end of clone CTA-282F2 is at 69682 in this sequence. The true right
end of clone CTA-415G2 is at 55167 in this sequence.

```
FEATURES             Location/Qualifiers
source          1..145880
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22"
                /clone="RP1-302D9"
                /clone_lib="RPCI-1"
repeat_region   188..245
                /note="MER3 repeat: matches 144..209 of consensus"
repeat_region   246..571
                /note="AluSx repeat: matches 1..312 of consensus"
repeat_region   572..759
                /note="MER3 repeat: matches 1..144 of consensus"
repeat_region   783..933
                /note="MER5A repeat: matches 26..187 of consensus"
repeat_region   1033..1336
                /note="AluSp repeat: matches 1..299 of consensus"
repeat_region   1450..1683
                /note="MIR repeat: matches 24..160 of consensus"
repeat_region   1687..1752
                /note="L2 repeat: matches 2593..2661 of consensus"
repeat_region   2350..2660
                /note="AluSc repeat: matches 3..309 of consensus"
repeat_region   2684..2981
                /note="AluSq repeat: matches 2..300 of consensus"
repeat_region   3323..3343
                /note="MLT1E repeat: matches 116..136 of consensus"
repeat_region   3344..3652
                /note="AluY repeat: matches 1..309 of consensus"
repeat_region   3653..3928
                /note="MLT1E repeat: matches 136..359 of consensus"
repeat_region   3929..4278
                /note="THE1B repeat: matches 3..364 of consensus"
repeat_region   4279..4485
                /note="MLT1E repeat: matches 359..568 of consensus"
repeat_region   5073..5176
                /note="52 copies 2 mer ct 78 conserved"
repeat_region   5181..5491
                /note="AluJb repeat: matches 1..311 of consensus"
repeat_region   6369..6485
                /note="L2 repeat: matches 2579..2705 of consensus"
repeat_region   6647..6685
                /note="MADE1 repeat: matches 1..23 of consensus"
repeat_region   6686..6987
                /note="AluSx repeat: matches 1..302 of consensus"
repeat_region   6988..7036
                /note="MADE1 repeat: matches 23..77 of consensus"
repeat_region   7482..7754
                /note="AluJb repeat: matches 9..290 of consensus"
repeat_region   7775..8060
                /note="AluJo repeat: matches 1..295 of consensus"
repeat_region   8414..8551
                /note="L2 repeat: matches 2553..2706 of consensus"
repeat_region   8914..9030
                /note="MIR repeat: matches 147..262 of consensus"
repeat_region   9110..9280
                /note="MIR repeat: matches 91..262 of consensus"
repeat_region   9283..9412
                /note="MIR repeat: matches 15..144 of consensus"
repeat_region   9521..9679
                /note="FAM repeat: matches 3..161 of consensus"
repeat_region   9820..10225
                /note="MSTB repeat: matches 2..425 of consensus"
misc_feature    complement(10179..10678)
                /note="match: GSS: Em:B56592"
misc_feature    complement(10204..10728)
                /note="match: GSS: Em:AQ701486"
misc_feature    complement(10249..10706)
                /note="match: GSS: Em:AQ225495"
repeat_region   10312..10383
                /note="MIR repeat: matches 79..150 of consensus"
misc_feature    10718..11310
                /note="match: GSS: Em:B14024"
misc_feature    10784..11201
                /note="match: GSS: Em:B43656"
repeat_region   11838..11946
                /note="MIR repeat: matches 20..137 of consensus"
repeat_region   12174..12445
                /note="L2 repeat: matches 1988..2275 of consensus"
repeat_region   12444..12642
                /note="AluSp repeat: matches 1..302 of consensus"
repeat_region   13017..13369
                /note="MIR repeat: matches 63..241 of consensus"
misc_feature    13331..13397
                /note="match: STS: Em:G49301"
repeat_region   13398..13698
                /note="MIR repeat: matches 174..244 of consensus"
repeat_region   13699..13810
                /note="AluSp repeat: matches 1..302 of consensus"
repeat_region   13806..13919
                /note="MIR repeat: matches 76..174 of consensus"
repeat_region   13945..14060
                /note="MIR repeat: matches 77..189 of consensus"
repeat_region   14061..14367
                /note="MIR repeat: matches 24..142 of consensus"
repeat_region   14368..14452
                /note="AluY repeat: matches 1..301 of consensus"
repeat_region   14589..14679
                /note="MIR repeat: matches 141..225 of consensus"
repeat_region   14597..15201
                /note="MIR repeat: matches 173..262 of consensus"
misc_feature    14616..15060
                /note="match: GSS: Em:AQ553482"
misc_feature    14868..15040
                /note="match: GSS: Em:AQ370601"
repeat_region   15071..15188
                /note="MIR repeat: matches 49..233 of consensus"
repeat_region   15304..15399
                /note="L2 repeat: matches 2112..2239 of consensus"
repeat_region   15490..15662
                /note="MLT1B repeat: matches 1..99 of consensus"
repeat_region   15669..15727
                /note="AluSq repeat: matches 2..114 of consensus"
repeat_region   15728..16027
                /note="MLT1B repeat: matches 119..178 of consensus"
repeat_region   16028..16245
                /note="AluSc repeat: matches 1..299 of consensus"
repeat_region   16546..16854
                /note="MLT1B repeat: matches 178..390 of consensus"
repeat_region   18296..18323
                /note="AluY repeat: matches 1..300 of consensus"
repeat_region   18324..18392
                /note="MSTA repeat: matches 2..29 of consensus"
```

```
repeat_region   /note="MER66-internal repeat: matches 4919..4993 of
18393..18712       consensus"
repeat_region   /note="AluJb repeat: matches 1..311 of consensus"
18713..19133
repeat_region   /note="MER66-internal repeat: matches 4548..4919 of
19133..19513       consensus"
misc_feature    /note="match: GSS: Em:AQ005063"
complement(18872..19230)
misc_feature    /note="match: GSS: Em:B14179"
19951..19719
repeat_region   /note="HERVFH21 repeat: matches 4657..5784 of consensus"
19537..20290
repeat_region   /note="33 copies 2 mer ta 68 conserved"
20317..20382
repeat_region   /note="77 copies 2 mer tt 70 conserved"
20513..20666
repeat_region   /note="AluSg1 repeat: matches 1..306 of consensus"
20682..21008
repeat_region   /note="HUERS-P3 repeat: matches 4410..4713 of consensus"
21239..21553
repeat_region   /note="THE1B repeat: matches 1..364 of consensus"
21682..22254
repeat_region   /note="MER66-internal repeat: matches 2186..2417 of
22302..22537       consensus"
repeat_region   /note="AluSp repeat: matches 1..313 of consensus"
22538..22850
repeat_region   /note="MER66-internal repeat: matches 1210..2186 of
22851..23801       consensus"
repeat_region   /note="MER66-internal repeat: matches 3017..3102 of
23905..23989
```

```
Query Match     750; Conservative 300; Mismatches 239; Indels 0; Gaps 0;
Best Local Similarity 58.2%; Pred. No. 7, 9e-215;
                Score 921; DB 9; Length 145880;

Qy     443  tnggncncncnggngcncnggngnytncccnacnatayacnggngaratrwsngaratgacna  502
            ||::|   |:|| ||:: |   |:||:: |  || ||::| :| ||::| : ||:||  |
Db   37254  TGGAATGNTCCTGGGTNCGTCCCCAGGTTNACCACAAATANTACAGGAGAAATAAGTGACACAA  37313

Qy     503  artgyccngayathgarngnwsngcnttyacngtnaaarytnwsnggnaarytnc  562
            ||::|::| :|  ||:|::||  |:||::||:| |::||  ||::|| ||||::|
Db   37314  AATGCCCCTGTCCTGATATATAGAAAAGGTCAGCCTTTACTGTGAAGGTCAGTGGAAAACTTC  37373

Qy     563  cnytnccnttyaarccnathathttyacnggngtnaytntatayaayrcncaarmgngayytna  622
            |::|::||::|  | ||:||::|||::||::||||| :| |:|::| :||| :||:|||:||
Db   37374  CTCTTCCTTTCAAGCCCATCATCTTCACAGGGGTCCTGTACAATGCCCAGAGGGATTTAA  37433

Qy     623  argargcnatgggngtnttygcntgymgngtnccnggnaaytaytayswswsnttygayg  682
            ||:|| :||:|||| ||||:||:| |||| ||:| |   |:|| |:|| :| || :|
Db   37434  AGGAGGCCATGGGAGTCTTTGCTTGCCTGGGAATTACTACTCCAGCTTTGATG  37493

Qy     683  tngarytncaycayctgyarrgtnaatgmgncaacarathytngcnaaya  742
            |:|| || |:|   | |  ||:| || |:| || ||::|| ||| |:|
Db   37494  TTGACTGCATCAATGCAAGGTGAATNATTTTGGCTNAATGAGGAAGCAAATNTACTGGTTA  37553

Qy     743  argargaratwsnaarcarwsnathcarygargtnacntgggtnytnaytnaargcnt  802
            ||:|| |:||::|  | |||::|  |||:| ||||:| ||::| |:||  ||  ||
Db   37554  AGGAGAAAATTTCTNAAGCACGAAAGCCATTCAAGAGGTGACTTGGCTGCGTTAAGGCAT  37613

Qy     803  tywsnttyathmgngargcngnarcayaarwsnsngaraytncaycncngayaaygtna  862
            |:|| ||::| |::| ||:| |  | |::|||::|::|||::|| |:||:||::|:|
Db   37614  TCAGTTCATNAAGGAGCNAGGCAGGCATAAGAGTTCAGAAAATTNGCACCCTGACAATGTGA  37673

Qy     863  thaararaaaraaycnttywsngarggnaartyaaryytngcngcngarathtgyatht  922
            ||::| || |::|::|||||::|:||::| |  |:|::|:|::|:|| |||||:|:|
Db   37674  TAAAAAAAGAAAAACCCCATTTTCTGAGGGGAAATNCAAGCTNGGCTNCCAGAAATTTGCATAT  37733

Qy     923  gyaaygargngarytnaaytcnacargayayggngaraayathwsntggacntgyc  982
            |:|| :||  | |  |::|::|| :|| | ||::| ||  |:||||::|| || ||
Db   37734  GTAATGAGGAGCTGAANTGTTAATCCTCAGACACAATGGGGAAAATATCTCTGGACATGTC  37793
```

```
Qy     983  armgnwsnwsncarcarwsnathaaarwsnytngcntggmgngccnmgmgmgnaartggttyt  1042
            ||:| :| |:|  | ||::|  |  ||::|||||:||:|::|::|::|:||  ||||:|||:
Db   37794  AGAGGTCTTCACAGCAGTCATCAAAATCACTGGCCTGGAGCCCTAGAGAAAATGGTTTT  37853

Qy    1043  gyggnacnggngccnsnytntgytgygtncarccnmgngayytngtnccntgyggtnc  1102
            |:||::| :||||:: :|:|||:|:|: | ||  |::|::||:||:|||:| ||:|||
Db   37854  GTGGGACNAGGGCCCAGGGTNCGTGCNTGTGTGCACTAGAGACTTGGTTGCCCTGTGTCC  37913

Qy    1103  cngtnaaywsnsngcngtngcnwsngargngngayggngcnwsnccnaarccntggacaryytyccncnwsng  1162
            ||::| || |:|::|::|||:| |:||:|:|:|| |::|:|::|::|::| |:|||| |::|::||:|::|
Db   37914  TTCAGAAGATATATGGAAAACCCTGGGTGCAAGACTCAGTGACTTGGTTGCCCTGTGTCC  37973

Qy    1163  gngtngarccncngtngngcngcnaaraarwsnmgnathgargtntggargarcccncnathmgnt  1222
            |::||:| |||:||:||::|:|::| | ||:|:|||:||:|::|||:||:| ||||:|:||:|||
Db   37974  GTGTTGAGCCCTGTGGGGTGCAAAGAAGTCAAGAAATTGAGGTTTGGAAGACTTCAAGATGAT  38033

Qy    1223  tycararathtaygngnaaycnntggatgtgcnmgncaraarttygcngtnggngtnggnw  1282
            |:||  | || ||:|:|   ||:||||:|||:| :|||  ||::|::|::||::|||::|
Db   38034  TTCAGAAGATATATGGAAAACCCCTGGGATGCCAGAGAAGTTTGCTGTAGGGGTTGTGGGGCT  38093

Qy    1283  snwsntggmgnacnwsngngtngtncaraaargnaaytngnngtngggarccncncnc  1342
            :| :||||:|:| :|::|::||||::| |  ||:|  ||:| :||:||:| || |:|:|
Db   38094  CCTCATCGAGAACCTCTCCAAGGGTAGTACAAAAGGGAAAATGTTGGGGTGGGAGCCCCCAC  38153

Qy    1343  aymgngtnccnwsngggngcccnwsnwsnmgmgncngtnmgmgnwsnccncnwsnwsnm  1402
            | :||:||:||::|:|:|:||:|::|::|:|:|:|::||:|:|:|::| ||:| |:| |
Db   38154  ACAGAGTCCCCAGTGGGGCTCCATCTAGTAGACTGGTGCGAGAGAAGTCCACATCCTCCA  38213

Qy    1403  gnytncararggnmgnwsnacngaywsnytncarcaygtnccngararaarwsnacngaya  1462
            |::||||   ||:|:|::| |:||::|::||| ||:| ||:||    |  ||::| |::|
Db   38214  GACTCCAGAAGGGTAGATCCACTGACAGCTTGGCATGTGCCTGAAAAATCCACAGACA  38273

Qy    1463  cncarcngycatccngtnaaargcngnggnatggaarwsngtnccntayaaarcngtntg  1522
            |:|| |::|:| :|:||:|  ||:|:|::||:|| |||::|||:|:| | :| :|:| ||
Db   38274  CTCAGTCGCCAGCCTGTGAAAAGCAGCAGCGGATGGAAGTCACCTACTACAAAAACCGTAGTGG  38333

Qy    1523  cngarytnacnaaracngtnggnathtayytnaytncaytgtgycaygayytngaygtnamgnc  1582
            ||:| ||:||  | |||:||:|:||||:|||:||:||| :| |:|::|:|::|||:|:| |:|
Db   38334  CAGAGCTGACCAAGACCGTGGGAATCTACCTCTTCGCATGTGTCATGACCGTGGACGTGAGAC  38393

Qy    1583  ayggngtnaaarmgngaycaytygygaytgygccnacnggngnttymgna  1642
            | ||::||||  |::|:| |::|:|:|| |:|::| |::||:|:|:|:| |
Db   38394  ATGGAGTCAAAAAGAGATCATTTTGGGAGCTGCCCCACTGGATTCGGA  38453

Qy    1643  cntayatgggngnccngtncccnytntgytyggncaarttyttycycncngayaaygtna  1702
            |:||  |||::|::|:||:||:| |:|:|:|||:|| :|||:||:| ||::|| |:||
Db   38454  CTTATATGGGGCCCGTACCCCTTTGTTTTGGCCAATTTTTTTCCATTTGGAACTGCCCTAT  38513

Qy    1703  tyacncaartgyytntayytncaytcaytgatg  1731
            |:|| |||:|::|:|||||:||| |:|::|:|
Db   38514  TTAACCCAATGCCTGTACCTCCATTGTATG  38542
```

```
REFERENCE
  AUTHORS   Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A. Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., HulyK,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Hohnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Muzny,D., Neal,J., Nelson,D., Newton,J.,
            Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E.,
            Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A.,
            Payton,B., Peery,D., Perez,L., Peters,L., Pickens,R., Primus,E.,
            Pu,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojukokan,I.,
            Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H.,
            Shim,C., Shooshtari,N. Sisson,I., Sodergren,E., Sonaike,T.,
            Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P.,
            Tamerisa,A., Taylor,H., Tang,H., Tansey,J., Taylor,C., Taylor,T.,
            Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V.,
            Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            WleczyK,R., Wooden,S., Worley,K., Wu,C., Wu,V., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R., Nelson,D. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 154090)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 154090)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 25, 2000 this sequence version replaced gi:9664948.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu
```

```
            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.
            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.
            QUALSTAT-REPORT---------
            -------- Summary Statistics --------
            Contig length:                                154090
            Phrap values in estimate:                     153545
            Average error rate (BCM-Phrap estimate):      0.00261755
            Fraction of Phrap values less than 40 :       0.0284086
            Number of consensus changing edits:           22
            Number of N's in consensus :                  0
            -------- Consensus changing edits --------
            Position   Original+Context          Edited+Context
            9347       aaacacagacc(n)tttttttt    aaacacagacc(t)tttttttt
            10135      attgcccct(n)taggaaaaga    attgcccct(t)taggaaaaga
            10353      ctgtcaccag(n)aaaagatcca   ctgtcaccag(t)aaaagatcca
            10417      agcaatgcag(n)ggctacagaa   agcaatgcag(t)ggctacagaa
            61959      cttactattt(n)tgtttattct   cttactatt(t)tgtttattct
            61960      ttactattn(n)tgtttattcta   ttactatt(t)tgtttattcta
            61961      ttactattnt(g)ttattcaa     ttactattg(t)ttattctaa
            61982      caacacacaag(n)taggtttgg   caacacacaag(t)aggtttggt
            61983      ntaggtttgg(n)ttactcttcc   ntaggtttgg(t)ttactcttcc
            61993      caattgaccc(n)ctggtcattc   caattgaccc(t)ctggtcattc
            62205      ntaggtttgg(n)ttactcttcc   ttaggtttgg(t)ttactcttcc
            62956      actgcaacct(n)tgcctcccag   actgcaacct(c)tgcctcccag
            63722      tttacatata(n)cagtacagta   tttacatata(a)cagtacagta
            90212      cctagaaaaa(n)gaacnttct    cctagaaaaa(t)gacttttct
            90216      gaaaaaagac(n)ttttcttttt   gaaaaaagac(t)ttttcttttt
            90228      tttcctttta(n)ttttctttta   tttcctttta(t)ttttctttta
            90236      tanttttttc(n)atcttttagg   tatttttttc(t)atcttttagg
            90443      accttagatg(n)ctcctcccag   accttagatg(a)ctcctcccag
            91405      cacagcttaa(n)caaggtaaga   cacagcttaa(a)caaggtaaga
            122290     aaaaaaagaa(n)gtaggagaaa   aaaaaaagaa(a)gtaggagaaa
            137090     acagagaaag(n)caaaacactc   acagagaaag(a)caaaacactc
            141632     tgaagccact(n)gaaaagtaat   tgaagccact(t)gaaaagtaat
            -------- Distribution of Quality < 40 Bases --------
                1000|
                 900|
            #    800|
            bases 700|
                 600|
                 500|                    *    *    *    *    *    *
                 400|               *    *    *    *    *    *    *
                 300|               *    *    *    *    *    *    *
                 200|          *    *    *    *    *    *    *    *
                 100|     *    *    *    *    *    *    *    *    *
                   0|*    *    *    *    *    *    *    *    *    *
                     5   10   15   20   25   30   35   40
                             Phrap Value Range
```

```
CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.
ANNOTATION OF FEATURES:
     STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
     Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
     Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25.3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.
```

```
FEATURES             Location/Qualifiers
                     Version: 1.01 qxfo.
```

```
source          1..154090
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="12"
                /clone="RP11-13C3"

repeat_region   54..288
                /rpt_family="MIR"
repeat_region   1020..1146
                /rpt_family="MER3"
repeat_region   complement(1248..1584)
                /rpt_family="MER46B"
repeat_region   complement(1624..1682)
                /rpt_family="MIR"
repeat_region   2630..2761
                /rpt_family="MIR"
repeat_region   4088..4216
                /rpt_family="GA-rich"
repeat_region   5233..5501
                /rpt_family="L2"
repeat_region   complement(5450..5749)
                /rpt_family="MER102"
repeat_region   6571..6616
                /rpt_family="Aluy"
repeat_region   complement(6678..7320)
                /rpt_family="(TG)n"
repeat_region   7419..7454
                /rpt_family="L1ME3"
repeat_region   complement(7741..7807)
                /rpt_family="AT_rich"
repeat_region   8113..8267
                /rpt_family="MER5A"
repeat_region   8385..8587
                /rpt_family="MER5A"
repeat_region   8589..8630
                /rpt_family="MIR"
repeat_region   8666..8695
                /rpt_family="L2"
                /rpt_family="(TTA)n"
```

```
Query Match         28.8%;   Score 497.8;   DB 9;   Length 154090;
Best Local Similarity   47.7%;   Pred. No. 1.6e-110;
Matches  488;  Conservative  216;  Mismatches  282;  Indels  36;  Gaps  2;

Qy      744 rgargarathwsnaarcacaraywsnathacargagtnacntgggtnytnaargcntt 803
            ::|||:|||||:|:||||::||||  |:||||||||||||::|:|||:| |||| |||
Db   101633 GGAAGAAAATTTCTAAGTAGTAGGAAAAGCATTAAAGAGGTGACTTAGAATGTGGTTAAAGGTATT 101692

Qy      804 ywsnttyathmgngargcngarcayaaarwsnwsngaaraayytncaycccngayaaavgtnat 863
            :||::||||||||:|||||||||::||||:||||||||:|||||||::||||:||||||:||||
Db   101693 CAGTTTATATAAGGGAGGCAGAGCATNAAACGTTTTGAAAAATTTGCAGCCTCACACAATGTGAT 101752

Qy      864 haaraaraaraaayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
            :||||:||:|||||:||||||:||||||||::|||||||:||:|:|||||||||::|:|
Db   101753 AGAAAAGAAAAAAACCCATTTTCTGAGGGAAATTTGAAGCNTGGCTGCCAGAAAATTTGCATTAAG 101812

Qy      924 yaaygargarytynaaygtnaaycnacacaragayayggngaaraaayathwsnctggacntgyca 983
            :|||:|||||:||:|||:||||:|:|||||||:|||||||||||:|||||:||||||||||:|||
Db   101813 TAACAAGGAGCCTAAYGTTAAYCCCAAGACAAATGGGAAAATGTTCTCCAAGGCATATCA 101872

Qy      984 rmgnwswsncarcaarwsnathaarwsnytngcntggmgnccnagnmgnaaartggtytytg 1043
            :||:|||||||||||:|||||||||:||||||||||:||||||||:||||||||||||:|:|
Db   101873 GAGTGTCTTCATGGGCACGGCCATCCACAGCCCGGAGCCCTAGGAGAAAATGGTTTTG 101932

Qy     1044 yggnacngngnccngnwsnynytntgytgygtncarccnmngayytngtnccntgyyt--- 1100
            :|||||:|:|:|||||||||||:|:|:|:|:|||||||:||||:||||||::||:|||
Db   101933 TAGGCCAGGCCCAGGAGCCCCATGCTGTCGTGCCTAGGGACTTGGTGCCCTGTGTATA 101992

Qy     1101 --------------------ncngtnaaywsngcngtungcnwsngar 1128
                                 ||:||||||:|||:|:|||:||:||:||||
Db   101993 CCAGCTGCTCCAGTGCTGGTGGGGCCAGCATAGAGACGTTGGGCGCTGGCTTCGGAG 102052

Qy     1129 ggngcnwsnccnaarccnttggacarytnccnwsnggngtngngncnaaraar 1188
            ||:|||||||||||:||||||||||:||:||:|||||:||:|:|:|:||||||
Db   102053 GGTGCAAGCCCCAAGCCTTGGCAATTTCCATGTGTTGTTGTGAGCCTGCACAGAAG 102112

Qy     1189 wsnmgnnathgargtntgggarccnnacnathmgnttycaraararthtayygnaayccntgg 1248
            :||:||:|||||||:|:||||||||:|:|||:||||||||:||||||::|||:|||||:|||
Db   102113 TCAAGAATTGAGGCTTGGGAGCCTCCACTAGATTTCAGAGAGTGTATGGAAATGCCTGG 102172

Qy     1249 atgccnmgncaraarttygcngtngggtnggngnwsnwsntggmgnacnwsngcnmgntn 1308
            |||:|||||||||||:|||||:||||||:||:|:|||||:|||:||||:||:||:||:|
Db   102173 ATGTCCAGGCAGAGAGTTTGCTTCAGGGTGGGGCCCCATGGAGAACCTCTCGTAGGGCA 102232

Qy     1309 gtncacaargngnaaygtnggnt                                     1368
            ||:||||||:|:|||:||||||
Db   102233 GTCCAGAAGGGAAAATGTCGGGTTGGACCCATACACACAGT                   102292

Qy     1369 wsnmgncngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngay 1428
            :||:||:|:||:||:||:|||||:||:||:||:||||:|||||||:||:||:|:||||
Db   102293 AGGGGAGCTGTGAGAAGAAGGGCCCACCAATCCTCCAGCACCCCAGAAATGGTAGAGNCCAC 102352

Qy     1429 wsnytncarcaygtnccngaraarwsnacngayacncaartgycarccngtnaargcngcn 1488
            :||:||||||||||:|||||||||:||||:|||:||||||:||||||:||||||:||:||
Db   102353 AGCTTGCACCAYGTACCTGAAAAGCCAACAAGCTCAACACGCCCATGAAAGCAGCC 102412

Qy     1489 ggnatggaarwsngtnccntayaaracngtngtngcngaryntnacnaaracngtngnath 1548
            ||:|||||||:||||:|:|||||:||:||||:|:|||||:||:|||:||||:||||:|||
Db   102413 AAGAGGGAGCCNGTACCCTGCAAAGCCACAGAAGCAGGCTTCCCAAGACATGGGAACC 102472

Qy     1549 tayytnytncaytgycayagayytngaygtnmgncayggngtnaaarmgnaaycaytgygn 1608
            |||:||:||||:|||:||||||:||||:||:||:||:||:||||||:||:|||||:|:|
Db   102473 CACCTCTTGCACCAGTGTGACCTAGATGTGAAACATCAGTGAAACATATTAGA 102532

Qy     1609 gcnytnmgntntygaytgyccnacnggnttymgnacntayatgggngcccngt-nccnytntg 1667
            ||:||:||:|:|:|||||:|||||:||:|||:||||:|||||||||||:|:||  ||:||:|
Db   102533 ACTTTGAGATTTGAACTACCCTTCTGGGTTTCAGAATGGGGCATATATAGGCCCCTTTG 102592

Qy     1668 yttyggncarttytytycnttyggnacngcngtnttyacncarrtgyyytntayytncayty 1727
            :||:|||||||||:|:|:||:||||||:||:||:|||||:|||:||:|::|||:||:||:|
Db   102593 TTTTGGGCCAATTTCTCCCATTTAGAACAGCCTGTATTTTACTCAATGCCTGTATCGCCATTG 102652

Qy     1728 ya 1729
            :|
Db   102653 CA 102654
```

```
RESULT   3
AC026107
LOCUS       AC026107           186660 bp    DNA     linear   HTG 06-FEB-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT
            SEQUENCE, 1 unordered pieces.
ACCESSION   AC026107
VERSION     AC026107.22  GI:13899178
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186660)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Alsbrooks,S.L., Bimage,K., Blankenburg,K., Bonnin,D.,
            Barbaria,J., Benton,J., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Bouck,J., Bowie,S., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Buhay,C., Carson,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Howell,J., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
```

```
COMMENT
REFERENCE   2  (bases 1 to 186660)
  AUTHORS   Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,D.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,I.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,U., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186660)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:13877175.
            -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center project name: HAMK
            Center clone name: RP11-307L11
            -------------- Summary Statistics
            Sequencing vector: Plasmid;
            Sequencing vector: M13;
            Chemistry: Dye-primer Bodipy: 18% of reads
            Chemistry: Dye-terminator Big Dye: 82% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 186630 bases at least Q40
            Consensus quality: 186659 bases at least Q30
            Consensus quality: 186660 bases at least Q20
            Estimated insert size: 187779; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 10.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
FEATURES             Location/Qualifiers
     source          1..186660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-307L11"
BASE COUNT    57089 a  37346 c  35962 g  56263 t
ORIGIN
```

```
            Query Match   28.8%; Score 497.8; DB 2;  Length 186660;
            Best Local Similarity  47.7%; Pred. No. 1.8e-110;
            Matches  488; Conservative  216; Mismatches  282;  Indels  36; Gaps  2;

Qy    744  rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt  803
           ||| || ||||:|| |||||||: |  |||||:|| |||| || ||||: |  ||  |||
Db  71882  GGAGAAATTTCTAAGTAGCAAAGCATTAAAGAGGTGACATTCAYCCNGAYGAYGTNAT  71941

Qy    804  ywsnttyathmgngargcngarcayaarwsnwsnqaraayytncaycyccngayaygtnat  863
           |||:||| ||| || || ||||| ||:||:||| || || |||:| |||| ||| || ||
Db  71942  CAGTTATATATAAGGAGAGACATRAACGTTTGGGAAAATTTGCAGGCTGACAAYGTGAT  72001

Qy    864  haaraaraaraayccnttywsngargrggnaarttyaarytngcncngaraththgyathtg  923
           |:|| || || ||| | |  :||   |||||||| | |||| :|: ||:|||| ||||||
Db  72002  AGAAAAGAAAAAACCATTTTCTGAGGGGAAATTTGAAGCTGCTGCAGAAAATTTGCATAAG  72061

Qy    924  yaaygargarytnaaygtnaayccncargayaayggngaraayatwsntggacntgyca  983
           ||| || |||| |||| || |||||||| ||||||||:|| ||:|| ||:|| || |||
Db  72062  TAACAAGGAGGCCTAAYGTTAAYCCCAAGACAATAGGGAAAATGTCTCCAAGGACATATCA  72121

Qy    984  rmgnawsnwsncarcarwsnathaaarwsnytngcntggmgnccnmgnmgnaartggttytg  1043
           | |:|:  :|  :  ||  :|| | |:::  : |:||   | ||:||||||||||:|||||
Db  72122  GAGGTCTTCATGGCAGCCCATCCATCACAGCCCCGGAGGCCTAGGAGAAAATGGTTTG  72181

Qy   1044  yggnacnggngnccnggnwsnytntgytgytgtncaarccnmgnqayytngtnccntgygt--- 1100
           | ||| :|| | |: |  :| || |||:|: ||| | |:|||::::|::|||:||| ||||
Db  72182  TAGGCCAGGCCCCAGGACCCCAGCTCGTGTCGCAGCCTAGGGACTTGGTTGCCCTGTGTATA  72241

Qy   1101  --------------------------nccngtnaaywsngcngtngcwsngar  1128
                                      ::: ||| |||| ||| ||| ||| ||
Db  72242  CCAGCTGCTGCAGTGTGGCTGAAAGGGCCAGCATNAGAGCTTGGCCGTGGCTTCGGAG  72301

Qy   1129  ggngcnwsncnaarccntggcaryytnccnwsngggntngarccngtngggncnaaraar  1188
           ||:||: :|  :  |  ||||| :||  | :|  |||:| || ||:| |  || | |:|:
Db  72302  GGTGCAAGCCCCCAAGCCTTGGCAAATTTCCATGTGCGTTGAGCCTGAGGGTGCACAGAAG  72361

Qy   1189  wsnmgnathgargtntgggarccnccnaathmgnttycaraaraathtayggnaaywsngar  1248
           ||:|| || || | |||| || | ||:|  :||||||| ||  |  : ||:|| || :|||
Db  72362  TCAAGAATTGAGGCTTGGGAGCCTCCACCTAGATTCAGATTTGGAAAATGCCTGG  72421

Qy   1249  atgccnmgncaraarttygcngtngngtnggngwsnwsntggmgnacnwsngcnmgngtn  1308
           ||| || :| :  :|||:| ||:| ||  ||||:| || |||| || ||:|||| |||:|
Db  72422  ATGTCCAGGCACAGAAGTTTGCTTCAGGGGTGGGCCCCCATGGAGAACCTCTCGTAGGGCA  72481

Qy   1309  gtncaraargghnaaygtngggntgggngarccncncayamgnccnwsngggngtncnnwsn  1368
           ||| | || ||  || |||| |  |||| ||  | |||| |||| |:| ||:| || ||||
Db  72482  GTGCAGAAGGAAAATGTGGGGTTGGGGATTGTATGGAAACCGGGGCCTACTGGGCCCCATCT  72541

Qy   1369  wsnmgngcngtnmgnmgnwsncncnccnwsnswsnmgnytncaraaargnmgnwsnacngay  1428
           ||:|| || ||  :|  :|  || :|  :| |:   || | :|  ||:|   || | |:||:
Db  72542  AGGGGACTGTGAGAACGGGGCCACCANCCTCCAGCACCCCAGAAATGGTAGAATCCACCAAC  72601

Qy   1429  wsnytncarcaygtnccngaaraarwsnacngayacncaartgycarccngtnaaargncgn  1488
           ||:|||| || ||| ||||||||:||:||:||:|  || || :||  ||:| ||||| |:|
Db  72602  ACTTTGACACCATGTACCCTGGGAAAAGCCACAGACTCAACACCCCATGAAAGCAGCC  72661

Qy   1489  ggnatggarwsgtnccntayaaracngtngtngcngarytnacnaaracngtngngnath  1548
           || ||| || |::| |:||| || |:|| || ||:| ||| ||  | |:|| |: ||| |
Db  72662  AAGAGGGAGCCTGTACCCTGCAAAGCCACAGAAGCAGAACCCAAGACCAATGGGAAC  72721

Qy   1549  tayytnytncaytgycaygayytngaygtnmgncaygggngtnaaramgngaycayttyggn 1608
           || :|| | ||:||:|| ||||:||:|| :|  || || :|||| | ::|||| ||||||:
Db  72722  CACCTCTTTGCATCAGTGTGACCTAGAYGTGAAACATGCAGTCAAAGGAGGATCATATTAGA  72781

Qy   1609  gcnytnmgntygaytggycccnacngngnttymgncacntayatgggngnncngt-ncnytntg 1667
           || :|| ||| ||| || |  |||:|:||:||| |||  ||| ||||:|:  ||| :|| :|||
Db  72782  ACTTTGAGATTTGCAACTACCCTTCTGGGTTTCAGACTTGCATAGCCCCTTTG  72841

Qy   1668  yttyggncaartyttyyccnttyggnacngcngtnttyacncartgyytntayytncaytg  1727
           | | ||| ||||:|| : :|||| ||| | |:|| ||:|| ||:|::||||||:||||:||
Db  72842  TTTTGGGCCAATTTCTCCCCAATTTAGAACAGCTGTTATTTACTCAATGCCTGTTATCGCCATTG  72901

Qy   1728  ya 1729
           ||
```

```
Db     72902 CA 72903
       :|
RESULT 4
LOCUS       AL138962/c
DEFINITION  AL138962
            Human DNA sequence from clone RP11-279N8 on chromosome 13, complete
            sequence.
ACCESSION   AL138962
VERSION     AL138962.16  GI:12597036
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124518)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 29, 2001 this sequence version replaced gi:12584354.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise stated: all
            sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-279N8 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-279N8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-279N8 is at 124518 in this
            sequence. The true left end of clone RP11-528D24 is at 114969 in
            this sequence. The true right end of clone RP11-214O11 is at 100 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1..124518
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-279N8"
                     /clone_lib="RPCI-11.1"
     misc_feature    112388..112496
                     /note="Big dye primer and dGTP big dye chemistries only."
BASE COUNT    39295 a   26385 c   24794 g   34044 t
ORIGIN
```

```
Qy      744  rgargarathwsnaarcarcaarwsnathcargargtnacntggtnytnaargcntt 803
             :||:||  |||::||  | |::: :|||||:|||||| ||  | ||: | :||:|
```

```
Query Match    28.58;   Score 493.6;  DB 9;  Length 124518;
Best Local Similarity   47.5%;  Pred. No. 1.7e-109;
Matches    486;  Conservative  216;  Mismatches  285;  Indels   36;  Gaps   2;
```

```
Db    66716  GGAAGACATTTCTAAGCAGAAAAAGCATTCAAGAGGTGACTTGGGTACTGTTAGAGGCATC 66657
```

```
Qy     804  ywsnttyathmngaargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
            |||:|| |:| ||:|||   ||| :|||  :|| | |||| |: :|| :|| :| | ||:
Db   66656  CAGTTTTATAAGGGAAAACAGACTTAAAAAGTTTGAAAAAGTTTGCAGCCTGACAATGTGAT 66597
```

```
Qy     864  haararaaraaaycccnttywsngaargngnaaaarttyaarytngncngcnaarthtgyathg 923
            |::||| :||||:| |  || :|||| : | | ::|:|||  |:||||| |||| ||| |:|
Db   66596  AGAAAAGAAAATCCCATTTTCTGAGGACAGGAGAGTATAAAAGTCAAGCTGGCTGCAGAAATTGCATAAG 66537
```

```
Qy     924  yaaygargaaytnaaygtnaayccncargacyagayggngaraaaythwsntggacntgyca 983
            |||||| ||| |:|| :|: ||::| || | |:|| ||| | ||:| |: :|:|| ||:|
Db   66536  TAACGAGGAGGCCAAAATGTTAAТCACCAAGACAAATGGAGGAAATGTCTCCAGGACATGTCA 66477
```

```
Qy     984  rmgnwswsnccarcaarwsnathaarwsnytngcntggmgnccnmgmg---naartggtt 1040
            | ||:|| : :|| :|  || :|| ||: | | ||:|| ||: |  |:|   |||:||:||
Db   66476  GAGGTCATCACAGCATCCCCTCCCCATCACACAGGCCCTTGAGAGCTTTAGGAGGAAAAATGGTT 66417
```

```
Qy    1041  ytgygygnpacngngnccnggnwsnytntgytgytgtncacaccnmngnayytngtncccntgygt 1100
            :|: ||| | |||| |  | :|| || :|:|:|:|:|: :|| | | :||| |::||::| ||
Db   66416  TCATGGAAGGGGGCACAGGGTCCCTGTGTGTGTGCCAGCCTAGGGACTTGGTGTCCTGCAT 66357
```

```
Qy    1101  nccngtnaayws------------------ncgngtngcnwsnga 1127
            || ||| || :                  : ||| |||| || |
Db   66356  CCCAGCTCGCTCAGCTGTGGCGTGAAAAGGGGCAAATGTGGAGTTCAGCGTGTGGCTTCAGA 66297
```

```
Qy    1128  rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtngggngcnaaraa 1187
            | ||| |||  | :||||:|| ||:|:|| :|| || :|:|| ||| :|| | || |||||
Db   66296  GGGTGCAGCCCTTAAGCCTTGGCAGCTTCCCACATGAGTTAAGGCCTGTGAGTGCACAGAA 66237
```

```
Qy    1188  rvsnsmngnathgargtntggarccnccnnathmgnttycararaarathtaygngnaayycntg 1247
            | |||: | | || |:| ||| | ||:|:|||:|| ||:| :|:|  |:||  ||| ||| ||:
Db   66236  ATCAAGAACTGGGGTTTGGGAACCTCTGTCTAGATTTCAGAGGATATTCAGAGTTAATAACCTG 66177
```

```
Qy    1248  gatgccnmgncarraarttygcngtngtnggnwsnwsntggmacnwsngcnmgngt 1307
            | |||: || ||| ||  :|| | :|:|| || |:|| :| || ||| | |::|
Db   66176  GATGCCCAGACAGAAGTTTGCTGTATGGGAGGGGTTCATGAAAAAACCTCTGCTAGGGC 66117
```

```
Qy    1308  ngtncaaraargnaaygtnggntgggarccnccncaymgngtnccnwsnggnngncccnws 1367
            | || ::||| || ||:|| ::|: ||| || |||:|| :|| :|| |||:| ||::|
Db   66116  AGTGTGGGGAGGGAAATGTGGGCCACTATCCTCCAGCACCTCTCACACAGAGTCCCTA 66057
```

```
Qy    1368  nvsnsmngngcngtnmgngnwsncccnccnwsnmgnytncararggnmgnwsnacnga 1427
            |:|||:| | | ::| : :|| || | ||: :| :| ||:|||| || ||  |:|| |
Db   66056  TAGTGGAGCTGTGAAAAGTGGGCGCCACTATCCTCCAGACCCCAGAATGGTAGAATCCACTGA 65997
```

```
Qy    1428  ywsnsnytncarcaygtnccngaraarwsnacngayacncaartgycarccngtnaaargcngc 1487
            :|| | | ||:|| ||:|| ||| ||||  |||| || |||:| | | || :||| | ||:| ||
Db   65996  CAGCTTGCCACCATGCACCTGGAAAAAGCCACACACTCAAATGCCAGCCTGTGAAAACGAGC 65937
```

```
Qy    1488  nggnatggarwsngtncntayaaracngtngcngarytnacnaaracngtnggnat 1547
            | ||||||| :|| || || |  | ||:|| |||  | || ||  | ||:|| | |
Db   65936  TGGGAGGGAGACTGTACCATGCAAGACCACAGGGCAGAGCCCCAAGAACCATGGAAC 65877
```

```
Qy    1548  htayytnytcaytgycgayytngaygtnggncayggntnaaaringngayaaytygg 1607
            | |||:| | ||||:| |||:|||: ||| |: :||| | :||:||:||||| |||
Db   65876  CCACCTCTTGCATCAGCATGGAGCATGGATCAGTGAGTCAAAGGAGGATCATTTTGG 65817
```

```
Qy    1608  ngcnytmngnttygaytgyccnacncggnttygmgnacntayatgggncngtnccnytntg 1667
            | | :| :| ||||: || || |||| :||:|  |:|| || ||:|| |||:| |:| ||
Db   65816  ACCTTTAAGGATTTGGACTTCCCGCTGGATTTCAGACTTGCATGGGCCTTAGCCCCCTTTG 65757
```

```
Qy    1668  yttyggncarttyttyccnttyggnacngcngtnttyacncaartgyytntayytncaytg 1727
            :|| || |:|:| :|| |||  | || |: |||  | | ||||:| |::| |: ||||:|
Db   65756  TTTTGGGCCAAATTCCCCCATTTGGAAATGGGTATATTTTACCCAATGCCTGTTACCCCCATTG 65697
```

```
Qy    1728  yat 1730
           | |
Db   65696  TAT 65694
```

```
RESULT 5
AC008799
```

```
LOCUS       AC008799             123070 bp    DNA     linear   HTG 18-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2061E19, WORKING DRAFT
            SEQUENCE, 8 ordered pieces.
ACCESSION   AC008799
VERSION     AC008799.4  GI:9256046
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7709316.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 651088
            Center clone name: CITB-H1_2061E19
            -----
            Summary Statistics
            Consensus quality: 116295 bases at least Q40
            Consensus quality: 121288 bases at least Q30
            Consensus quality: 122086 bases at least Q20
            Estimated insert size: 123000; pulse field gel estimation
            Estimated insert size: 122770; sum-of-contigs estimation
            Quality coverage:    6.04 in Q20 bases; pulse field gel estimation
            Quality coverage:    6.05 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 8 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces
            is believed to be correct as given, however the sizes
            of the gaps between them are based on estimates that have
            provided by the submitter.
            This sequence will be replaced
            by the finished sequence as soon as it is available and
            the accession number will be preserved.
            1             9436: contig of 9436 bp in length
            9437          9536: gap of unknown length
            9537         26537: contig of 17001 bp in length
            26538        26637: gap of unknown length
            26638        45958: contig of 19321 bp in length
            45959        46058: gap of unknown length
            46059        68657: contig of 22599 bp in length
            68658        68757: gap of unknown length
            68758        72152: contig of 3395 bp in length
            72153        72252: gap of unknown length
            72253        81069: contig of 8817 bp in length
            81070        81169: gap of unknown length
            81170       121547: contig of 40378 bp in length
            121548      121647: gap of unknown length
            121648      123070: contig of 1423 bp in length
FEATURES             Location/Qualifiers
     source          1..123070
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2061E19"
                     /clone_lib="Caltech human BAC library D"
BASE COUNT    37763 a  24423 c  23855 g  36528 t    701 others
ORIGIN
```

Query Match   28.5%; Score 493; DB 2; Length 123070;

Best Local Similarity 46.7%; Pred. No. 2.3e-109;
Matches 495; Conservative 220; Mismatches 309; Indels 36; Gaps 2;

```
Qy        706 aayathcggytnaatgmgnaarcarathytngcnaaygargarathwsnaarcarcar 765
              ||:|  ||  |:|  :| |||  :|: |||:| ||| | :|| || |:|:|||:||||:
Db     107907 AACTTTTAACTTGAAAGAAAATGATTTTAGGGTATCTGCGAGAGAAAATTTCTAAGCAGCAA 107966

Qy        766 wsnathcaargartgnacntgggtnytnytnaargcnttywsnttyathmgnaargcngar 825
              |||:  ||  ||||||| :|| ||| | :|: || : |:| | ||:|  || || :|||
Db     107967 AGCATTCCAGAGGTGACTTGGGTCGTCGTTAAAGGTATTCAGTTTATAAGGAAGCAGAG 108026

Qy        826 cayaaarwsnwsngaraaaytncaycyconga ayaygtnathaaraaraaayccnttywsn 885
              | ||| | ||| |||| ||||: || | | |||||  ||| :|:|||||||| :|: :| |
Db     108027 CATAAAAAGTCTGAAAAATTGCAGTCCGACAATGGTGATAGAAAAGAAAATTCCATTTTCT 108086

Qy        886 gargaynaarttyaarytngcngcngarathtgyathcgyaaygargarytnaaygtnaay 945
              || ||| ||  | :|  ||||| |: ||:|  | | ||| | ||| || ||| ||:|| ||
Db     108087 GAGGAGAAAATTCAAGCTGTGGGCTGCAGAAAGTGTCCATAAGTAACAAGGAGCTGAATATTAAT 108146

Qy        946 cncaargayaaygngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 1005
              |||||| | | || ||| || |||||| | || ||: |:|| ||| | :|| ||  |
Db     108147 CCCAAAGACAATGGGGAAAACATCTCCAAGGACATGTCAGAGGTCTTCACGGCAGCCCTC 108206

Qy       1006 aarwsnnytngcntggmgnccncnmgnmg--naartggttytygggnacnggngccnggnwsn 1062
              :::|  :| |||||| ||:| | |:|||||  || |||:||||||| ||| |  |||:|
Db     108207 CCATCATAAACCCAAAGCCTAGCAGGAGAAAAATGGTTTGTTTGTGGGCCAGGCCCCAGGGTCC 108266

Qy       1063 ytntgytgygtgtncarccnmgnaayytngtncctgygtncngtnaa---------- 1109
              |:||| |:| ||: || |  ||: :| |||:|:|||| ||: |:|||
Db     108267 TTGTGCTGTGTCAGCTTGGTGTGCCTCGCAATCCCACCTGCTTCCACCATGGCT 108326

Qy       1110 ------------ywsngcngtngcnwsngargrggngcnwsnccnaarccntgg 1149
                          :| | ||| :| || ||||| |:|| :||::|  || |||
Db     108327 GAAAGGGGCCACATAGAGCTCGGGCGTCTGGCTTCAGAGGGCCAGGCCCAGGGTCC 108386

Qy       1150 carytnccnwsnggngtngarccngtngggcnaaaraarwsnmgnathgargtntgggar 1209
              || |||:|:|| ||||||| || |||||| |||| :||||| :| || |:|| :|||||
Db     108387 CACCTTCCATGTGTGTTGAGCCTGCAAGTGCACAGAGTCAACATTGGGGTTTGGGGAA 108446

Qy       1210 cnccnathmgnttycaraaratthtaygnnaayccntggatgcnnmgncaraartycgcn 1269
              | |||| ||:| |:|| ||||| ||| | |   | || | ||||||||| :|:||:|||
Db     108447 CCTCCACCTAGATTGCAGATGATGTATCGGAAAATGCCCGGATGACCACCAGCAGAAGTTTGCT 108506

Qy       1270 gtnggngtnggnwsnwsntggmgnacnwsngcnmgntgntgcaraargggnaaygtnggn 1329
              ||| ||||| || ||| ||| || |  |  |:| | |||:||||||||||||:||:||
Db     108507 GCAGGGGGCCAGGGCCTTCATGGAGACAACCTCTGCTAGGGAGAGTGCGAGAGAAAAATGTAGG 108566

Qy       1330 tgggargccncnccnaaytmgnwsnggtnccnwsnwsnmgncngtnmgmgmtwsn 1389
              |||| |   |  ||:|| |:|  |:  || | ||||| || | || || ||:|
Db     108567 TCAGAGCCCACACACAGAGTACCCACTAGAGCACTGCCTAGTGGAGCCTGAGAGAAGG 108626

Qy       1390 cnccnwsnwsmngnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1449
              |||| :| :|:| |:| || ||||| |:| |:  ||  ||| | || ||| || ||
Db     108627 CCAACATCCTCAGACTCCAGAATGGTAGGTCCGCCAACAGCTTGCACTGTGTTACCTGGA 108686

Qy       1450 aarwsnacngayacncaartgycarccngtnaargcngcngnnatggarwsngtnccntay 1509
              ||: :|| |  | |  ||||| |||| || ||||||| |  |  ||| |:|||:|| |
Db     108687 AAAGCCACAGGACCACTCATTGCCAGCCTGTGAAAGCAGCAGGAGGCTAACCCCTGC 108746

Qy       1510 aaracngtngtncgngaytnacnaaracngtnggnnathtayytnytcaytgycaygay 1569
              ||: |  | ||| | ||:  ||| |||  | |||  | ||: | | :||| |  |:|
Db     108747 AAAGTCACAGGGCGGCGGAGCTGTCCAAGACCTTGGGAACCTACCTCTTGCCTCAGCGTGAC 108806

Qy       1570 ytngaygtngmncayggngtnaarmgngaycaytnmgnttygaytgycn 1629
              :  |:||| |||| ||| |  |||  | || ||:||  | :|| | || |
Db     108807 CCGGATGTGAGACATGGAGTCAAAGGAGAATCATTTTGAGCTTTAGAGTTTGACTGCCT 108866

Qy       1630 acnggnttymgnacntaygtggncngtnccnytntgytyggncarttytyccntty 1689
              |  || ||: |  |||  | ||  | ||||| |:||| |||||| ||| :| | |
Db     108867 GCTGGAGTTTCAGACTGCATGGGGCGGCTGGTGAGCCCCTTTTGGCTTTTGGCCAATTTTTCCCCATTT 108926

Qy       1690 ggnacngcngtnttyacncaartgyytntayytncaytgya 1729
              || ||  || | |  ||| | |  |:| |:||| ||:|:|
```

```
Db 108927 GGAACAGCTGTATTACCCCATGCCTGTACCCGATTGTA 108966

RESULT    6
LOCUS       AC044889/c
ACCESSION   AC044889
DEFINITION  Homo sapiens chromosome 2 clone RP11-792C1 map 2, WORKING DRAFT
            SEQUENCE, 36 unordered pieces.
ACCESSION   AC044889
VERSION     AC044889.2  GI:8016676
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-792C1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Piasni,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirelli,A., Travers,M., Triglilo,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 22, 2000 this sequence version replaced gi:7543856.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------------------- Project Information
                Center project name: L9594
                Center clone name: 792_C1
            -------------------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 168749 bases at least Q40
                Consensus quality: 180047 bases at least Q30
                Consensus quality: 184014 bases at least Q20
                Insert size: 190000; agarose-fp
                Insert size: 186268; sum-of-contigs
                Quality coverage: 3.8 in Q20 bases; agarose-fp
                Quality coverage: 3.9 in Q20 bases; sum-of-contigs
            ---------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 36 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
FEATURES          Location/Qualifiers
     *      158014   189768: contig of 31755 bp in length.
source         1..189768
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="2"
               /map="2"
misc_feature   1..1015
               /clone="RP11-792C1"
               /clone_lib="RPCI-11 Human Male BAC"
misc_feature   1116..2284
               /note="assembly_fragment"
misc_feature   2385..3898
               /note="assembly_fragment"
misc_feature   3999..5259
               /note="assembly_fragment"
misc_feature   5360..6912
               /note="assembly_fragment"
misc_feature   7013..8145
               /note="assembly_fragment"
misc_feature   8246..9694
               /note="assembly_fragment"
misc_feature   9795..11004
               /note="assembly_fragment"
misc_feature   11105..12907
               /note="assembly_fragment"
misc_feature   13008..14916
               /note="assembly_fragment"
misc_feature   15017..17114
               /note="assembly_fragment"
misc_feature   17215..18527
               /note="assembly_fragment"
misc_feature   18628..20561
               /note="assembly_fragment"
misc_feature   20662..22484
               /note="assembly_fragment"
misc_feature   22585..25287
               /note="assembly_fragment"
misc_feature   25388..27603
               /note="assembly_fragment"
misc_feature   27704..30020
               /note="assembly_fragment"
misc_feature   3021..33472
               /note="assembly_fragment"
misc_feature   33573..38084
               /note="assembly_fragment"
misc_feature   38185..41069
               /note="assembly_fragment"
misc_feature   41170..45266
               /note="assembly_fragment"
misc_feature   45367..51092
               /note="assembly_fragment
               clone_end:T7
               vector_side:right"
misc_feature   51193..55262
               /note="assembly_fragment"
misc_feature   55363..60982
               /note="assembly_fragment"
```

```
Query Match    28.5%;  Score 493;  DB 2;  Length 189768;
Best Local Similarity  46.7%;  Pred. No. 2.7e-109;
     Matches  495;  Conservative  220;  Mismatches  309;  Indels  36;  Gaps  2;

Qy    706  aayathtggytnatgmgnaarcarathytngcnaayaaargarathwsnaarcarcar  765
           ||:|:||| || | |||| | | || |||:|||: ||||||
Db  86595  AACTTTTAACTTGAAAGAAATGAATTTAGGGTATCTGGCAGAAGAAAATTTCAAGCAGCA  86536

Qy    766  wsnathcaarargtnacntgggtnytnaytnaargcnttysnttyathmngargcngar  825
           ::||:|||::| ||:| |||:|| |||| :|||:| || |:|| ||||:||:|
Db  86535  AGCATTCCAGAGCGTGACTTGGGTGCTGTGTTAAAGGTATTCAGTTTTATAAGGAAGCAGAG  86476

Qy    826  cayaaarwswsnsnganraaaytncayccngayaaygyynaythahaaraaaraaaaycccntywsn  885
```

```
Db  86475  CATAAAAAGTCTGGAAAAATTTGCAGTCCGACAAGTGATAGAGAAAAAGAAAATTCATTTTCT  86416

Db  86415  GAGGAGAAAATTCAAGCTGGCTGCAGAAAATGTGCATAAAGAGTCTTGCAGGGTCC  86355

Qy    886  gargngnaartttyaaarytngcngcngnaarathcgyathtgyaayggargaryytnaaygtnaay  945
           |||:| :| | | ||||::| |: ||| | |:|| ||  ||: || | |:|| | || |
Db  86355  CCCAAGACAATGGGGAAAAACATCTCCAAGGCACATGTCAGAGGTCTTCACGGACAGCCCTC  86296

Qy    946  ccncargayaaygygngaaraayathwsngacntgycarmgnwswsncarcarwsnath  1005
           ||| ||| || || | |:||| |||| | | || || || | ::| |:||
Db  86295  CCATCATAAACCCAAAGGCCTTGAGGAGAAAAAGTGGTTTTGTGGGAACGGCCAGGGTCC  86236

Qy   1006  aarwsnaytngcntggngncccmgmng---naartggttytgyggnacngccnggnwsn  1062
           | |: |:|| | ||  |:||: ||  |||:||| | |||:|| | || | || ||
Db  86235  TTGTGCTGTGTGTGCAGCTTAGGGACTTGGTGCCCTGCCATCCCAGGCCATGGCT  86176

Qy   1063  ytntgytgtygtncarccnmgnggayytngtnccntgyytncngtaa----------  1109
           :|| |:|| | |||:|||| | ||| :|| ||:||| | ||:|| ||
Db  86175  GAAAGGGGCCAACATAGAGTCTGGGCTGTTGCTTTCAGAAGGTGCAAGCCCCAAGCCTTGG  86116

Qy   1110  ----------ywsngcngtngcnwsngargngngcnwsnccnaarccntgg  1149
                     |||:|| | | | | || || | | |: || ||
Db  86115  GAAAGGGGCCAACATAGAGTCTGGGCTGTTGCTTTCAGAAGGTGCAAGCCCCAAGCCTTGG  86056

Qy   1150  carytnccnwsngggngtngarccngtnggngcnaaraaarwsnmgnaathgargtntggar  1209
           ||: | |||| ||| ||||:||||| |||||| |: || ||:|||:|| |: || |||
Db  86055  CAGCTTCCAGTGGTGTGTTGAGCCTGCAAGTGCACAAAATTGGGGTTTGGGAA  86056

Qy   1210  ccnccnaathmgnttycaraatathtaygnaayccntggatgnccmgncaraaartygcn  1269
           ||| ||| ||:|||| |:| | ::|: | | || :|| ||| || | :| || |:||
Db  86055  CCTCCACCTAGATTGCAGATGATGTAGGAAAATGCCCGGATGACCACGGCAGAAGTTTGCT  85996

Qy   1270  gtngngngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraaargngaaygtnggn  1329
           ||| | || || || ||| ||:|| ||| ||:|| || ||:|| |:| |:|| || ||
Db  85995  GCAGGGGGGCCAACATAGAGTCGCCTTCATGGAGAACCTCTGCTAGGGCAGTCGAAGAAAATGTAGGG  85936

Qy   1330  tgggarcccncacaygngtnccnwsngggngcncccnwsnwsnmgncgngtnmgnmgnwsn  1389
           |||| || || ||||| ||||| | || ||| | || |||| || |: ||  | | ||
Db  85935  TCAGAGCCCACACACAGAGTACCCACTAGAGCACTGCCTAGTGGAGCTGTGTGAGAGAAGG  85876

Qy   1390  ccnccnwsnwsnmgnmgnvtncaraaargggmgnwsnacngaywswsnytncarcaygtncngar  1449
           ||| || || ||:||:| |||:| || |||| ||| || || || ||:|| |:||| | |||
Db  85875  CCACCAACCTCCAGACTTCCAGAATTGGTAGGTCCACAAACAGCTTGCACTGTGTACCTGGA  85816

Qy   1450  aarwsnacngayacnccaartgycarccngtnaargcngcngnnatggarwsngtncccntay  1509
           || || || ||| |  | |||||| || || ||:|| ||:| | || ||:|| |||| |:|
Db  85815  AAAGCCACCAGGCACTCATTGCCAGCCGTGTGAAAGCCACCAGGGCGTAGGGCCTATCCCCTGC  85756

Qy   1510  aaracngtngtncngaryytnacnaaracncngtnggnnathtayytnytnytcaytgycyy  1569
           || ||||:||||:|| |:|:||| || |: ||:|| ||| | |||:|::|:||| | |:|
Db  85755  AAAGTCACAGGGGCGGGCGAAGCTGTTCCAAACAGCCTGGGAACCTACCTCTTGCCCTCAGCGTGAC  85696

Qy   1570  ytngaygytnmgncayggngtnaarmgngaycayytyggngcncytnmgnttygaytgycn  1629
           :||:|:||| ||| || ||:||:|| || | | ||:|| | | |:|: ||:|| | ||| |:
Db  85695  CCGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGGAGTTTAAGATTTGACTCTGCCCT  85636

Qy   1630  acnggnttymgnacntayatgyggngccnggncctntcnytygggncarttytyycntty  1689
           || |||||:||| || ||:|  ||: || || |:| | ||:||| || ||:|::|:||
Db  85635  GCTGGATTTCAGACTGCATGGGGCCTGTAGCCCCTTTGCTTTGGCCAAATTTCTCCCATTT  85576

Qy   1690  ggnacncngtnttyaccncaartgyytntayytncaytgya  1729
           ||:|| |:||:|| ||| ||:|| |:|| ||||| |:|
Db  85575  GGAACAGCTGTATTTACCCATGCCCGTACCCGATTGTA  85536
```

```
RESULT 7
AC017063

LOCUS         AC017063   198545 bp   DNA   linear   HTG   02-NOV-2001
DEFINITION    Homo sapiens chromosome 4 clone RP11-354H17, WORKING DRAFT
              SEQUENCE, 4 unordered pieces.
ACCESSION     AC017063
VERSION       AC017063.7   GI:16596637
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE        human.
```

```
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198545)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198545)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 2, 2001 this sequence version replaced gi:15741601.
            --------------- Genome Center ---------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ------------ Project Information ------------
            Center project name: H_NH0354H17
            --------------- Summary Statistics ---------------
            Sequencing vector: M13; 35%
            Sequencing vector: plasmid; 65%
            Chemistry: Dye-primer ET; 34% of reads
            Chemistry: Dye-terminator Big Dye; 66% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197437 bases at least Q40
            Consensus quality: 198145 bases at least Q30
            Consensus quality: 198648 bases at least Q20
            Insert size: 200000; agarose-fp
            Insert size: 200801; sum-of-contigs
            Quality coverage: 10.58 in Q20 bases; agarose-fp
            Quality coverage: 10.55 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1    1174: contig of 1174 bp in length
            *     1175    1274: gap of unknown length
            *     1275    2716: contig of 1442 bp in length
            *     2717    2816: gap of unknown length
            *     2817    4952: contig of 2136 bp in length
            *     4953    5052: gap of unknown length
            *     5053  198545: contig of 193493 bp in length.
FEATURES             Location/Qualifiers
     source          1..198545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-354H17"
     misc_feature    1..1174
                     /note="assembly_name:Contig24"
     misc_feature    1275..2716
                     /note="assembly_name:Contig38"
     misc_feature    2817..4952
                     /note="assembly_name:Contig42"
     misc_feature    5053..198545
                     /note="assembly_name:Contig44"
                     vector_side:right"
                     clone_end:SP6
BASE COUNT     62963 a  36710 c  37248 g  61324 t     300 others
ORIGIN
```

Query Match   28.3%;   Score 490.4;   DB 2;   Length 198545;

```
Best Local Similarity   47.0%;  Pred. No. 1.2e-108;
Matches  498;  Conservative 213;  Mismatches 314;  Indels  34;  Gaps  2;

Qy    706 aayathtggytnaatngnaarcarathytngcnaayaargargarathwsnaarcarcar  765
          |!:||:||| ||:|||||:|||||||:|||| |:
Db  83103 AACTTTTAACTTGGGACAGAGATCAATTTAGGGTATCTGCAGATAAATTTCTAAGCAGCAA  83162

Qy    766 wsnathcargargtnacntgggtnytnaargcnttywsnttyathmgngaargcngar  825
          ::||| || |||||||||||||||| || ||||||||||||||:|||||:|||| ||:
Db  83163 AGCATTCAAGCACGTGACTTGGGTGCGTGTTAAAGGGCATTCAGTTTTATAAGGGAAGCAGAG  83222

Qy    826 cayaarwsnwsngaraayytncaaccngayaaygaygtnathaaraaaraaraayccnttywsn  885
          |!:||: |:|| |||:|||||||||||||:|||||||||| ||:|||||||||||||||:||
Db  83223 CATAAAAAATTTGGAAAAATTTGCAGCCTGACAAATGTGATAGAAAAGAAAAAACCCATTTTCT  83282

Qy    886 garggnaarttyaarytngcngcnagarathtgyathtgyaaygargarytnaaaygtnaay  945
          |!:|| ||||||||||:|||:| |:|||||||| ||||||||||||||||:|||
Db  83283 GAGGAGAAAATTGAAGCTGGCTGCAGAAAATTTACATAAGTAATGAGGAGACCTGAATGTTAAT  83342

Qy   1006 aarwsnytngcntgggmgnccnmgnmgnaaartggttytygyggnaacngngncnnggnwsnytn  1065
          :::|  || || || || | ||||||| | |||||| ||||||| ||||||||:||||||||:
Db  83403 CCATCACAGGCCGTGAAGGCCTAGGAGAGAAAAATGGTTTTCGGGCGCCCACTCCAGGGTCCCTG  83462

Qy   1066 tgytgygtncarccnmgngayytngtncccntgygtnccngtnaayws-----------  1112
          |! |||||||||||||||||| ||||| | || || | ||||||
Db  83463 TGCTGTGTGCAGTCTAGGTACTTGGGTCCCCTGCATCCTAGCCACTGACATGACTAAA  83522

Qy   1113 ---------------ngcngtngcnwsngargangngcnwsnccnaarccnctnggcar  1152
                         :||:||||||||||| ||||||||||||||||||:|||:
Db  83523 AGGGGCCCAAGGTACAGCCTCAGGCTGCTGTTGCTTCAGAGGGGTGGAAGCCTGCAAGCCCTGGCGG  83582

Qy   1153 ytnccnwsnggngtngarccncngtnggngcnaaaarwsnmgnaathgargtntgggarccn  1212
          |::|:| | ||| |||  ::| ||||||||||| | |||| |||| |:||||||||
Db  83583 CTTCCAAATGGTATTGAACCTGCGGGTTCACAGAAGTCAAGAATTAAGGTTTGGGGAACCT  83642

Qy   1213 ccnathmgnttycaraaraathtayggnaaayccntggatgcnmgncaaraarttygcngtn  1272
          ||:||||||||||||| ||||||||||||||||||||||| |||:||
Db  83643 CCGCCTAGATTTCAGAAAAATGTATGGAAAATGCCTAGCCCAGGCGAGAATTTTGCTGCA  83702

Qy   1273 ggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraccaygtnccngaaraar  1332
          |:||| || |||:|||||||||||||||| |  |||:|||||||||| ||| ::|
Db  83703 GGAGTGGGGCCCTCATGGAGAACCTCTCGCTAGGCAGTGCAGAAGGGAAAAGTGAGGTCA  83762

Qy   1393 ccnwsnwsnmgnytncaraargnmgnwsnacngaywsnytncarcayngtnccngaraar  1452
          || ||:|||| ||:| |||||||| ::|||||||||||||||||||||||:|| ||||
Db  83823 CTGTCCTCCAGACCCCAGAAATGGTAGATCCCACTGAGAGCTTCCACCATGCACTTGGAAAA  83882

Qy   1453 wsnacngayacncaartgycatccngtnaargcngcnggnaatggaarwsngtnccntayaar  1512
          :||:|| |||| ||||| || | ||:|:|||||||||||||||| |:|:|||| |::|||
Db  83883 GCCACAGACACTGAACACCGACATGAAAAGCAGCCAGGGAGGGGTGGCTGTACCCTACAAA  83942

Qy   1513 acngtngtngcngarytnacncaaracngtnggngnathtayytnaytytcayaytygcaaygayytn  1572
          |:|||||| ||||| | ::| || :|||||||| ||||||||| ||||:|||||| ||
Db  83943 GCCACAGGGGTGAAGTTGGCCAAGACCACAGGAACCCACTCTTGCATCACAGTACTTG  84002

Qy   1573 gaygtcnmgncayggngntnaarmgngaycaytydtggngcnytcnmgntygtgaytygtyccnacn  1632
          ||||:| || |||||||||:||||||||| :|:| |:||||||||||||||:||||||||:||||
Db  84003 GATGTGAGACATGGTGTCAAAGGACACTATATTTGGAGCTTTAAGATTTGACTGCCCACT  84062

Qy   1633 ggnttymgnacntayatgggnccngt-nccnytntcgytygggncarttytyccnttygg  1691
          |||||:||:||||| ||||||||| | |:|||:|:| ::|||||||| :|::|::|||||
Db  84003 GGATTTTGGGACTTGCACGGGCGCTTGTAAACCCCTTGTTTGGGCCCAAATGTCTCTCATTTGG  84122

Qy   1692 nacngncngtnttyacncaartgyytntayytncaaytgyat  1730
          |!||||||||||:||::| |||::|||| ||::|:| |
```

```
Db   84123 ATTTGCTGTATTTATCCAATGCCTGTACTGCCAATGTTT 84161

RESULT   8

AC025233/c
AC025233

LOCUS       AC025233              161575 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-333E1, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC025233
VERSION     AC025233.4  GI:8954317
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161575)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161575)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:7801489.
            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            -------- Project Information --------
            Center project name: H_NH0333E01
            -------- Summary Statistics --------
            Sequencing vector: M13; 90%
            Sequencing vector: plasmid; 10%
            Chemistry: Dye-primer ET; 81% of reads
            Chemistry: Dye-terminator Big Dye; 19% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 144558 bases at least Q40
            Consensus quality: 149743 bases at least Q30
            Consensus quality: 152625 bases at least Q20
            Insert size: 176000; agarose-fp
            Insert size: 158975; sum-of-contigs
            Quality coverage: 3.48 in Q20 bases; agarose-fp
            Quality coverage: 3.75 in Q20 bases; sum-of-contigs
            --------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1      1141: contig of 1141 bp in length
            *     1142      1241: gap of unknown length
            *     1242      2424: contig of 1183 bp in length
            *     2425      2524: gap of unknown length
            *     2525      3961: contig of 1437 bp in length
            *     3962      4061: gap of unknown length
            *     4062      5712: contig of 1651 bp in length
            *     5713      5812: gap of unknown length
            *     5813      8019: contig of 2207 bp in length
            *     8020      8119: gap of unknown length
            *     8120     10208: contig of 2089 bp in length
            *    10209     10308: gap of unknown length
            *    10309     11734: contig of 1426 bp in length
            *    11735     11834: gap of unknown length
            *    11835     13265: contig of 1431 bp in length
            *    13266     13365: gap of unknown length
            *    13366     14425: contig of 1060 bp in length
            *    14426     14525: gap of unknown length
            *    14526     15609: contig of 1084 bp in length
            *    15610     15709: gap of unknown length
            *    15710     17529: contig of 1820 bp in length
            *    17530     17629: gap of unknown length
            *    17630     20351: contig of 2722 bp in length
            *    20352     20451: gap of unknown length
            *    20452     23753: contig of 3302 bp in length
            *    23754     23853: gap of unknown length
            *    23854     26537: contig of 2684 bp in length
            *    26538     26637: gap of unknown length
            *    26638     31339: contig of 4702 bp in length
            *    31340     31439: gap of unknown length
            *    31440     37041: contig of 5602 bp in length
            *    37042     37141: gap of unknown length
            *    37142     45268: contig of 8127 bp in length
            *    45269     45368: gap of unknown length
            *    45369     52323: contig of 6955 bp in length
            *    52324     52423: gap of unknown length
            *    52424     60239: contig of 7816 bp in length
            *    60240     60339: gap of unknown length
            *    60340     68385: contig of 8046 bp in length
            *    68386     68485: gap of unknown length
            *    68486     78834: contig of 10349 bp in length
            *    78835     78934: gap of unknown length
            *    78935     90422: contig of 11488 bp in length
            *    90423     90522: gap of unknown length
            *    90523    100763: contig of 10241 bp in length
            *   100764    100863: gap of unknown length
            *   100864    111672: contig of 10809 bp in length
            *   111673    111772: gap of unknown length
            *   111773    126855: contig of 15083 bp in length
            *   126856    126955: gap of unknown length
            *   126956    142557: contig of 15602 bp in length
            *   142558    142657: gap of unknown length
            *   142658    161575: contig of 18918 bp in length.

FEATURES             Location/Qualifiers
     source          1..161575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-333E1"
BASE COUNT    44288 a   35582 c   35114 g   43686 t   2605 others
ORIGIN

Query Match     28.2%;  Score 488.8;  DB 2;  Length 161575;
Best Local Similarity   48.5%;
Matches   485;  Conservative  202;  Mismatches  280;  Indels  34;  Gaps   2;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytngnytnaargcntt 803
          :|||  |:|||:|| | :|| |: ||| |:|:|||.||:::: ||| ||:|||:|||:|:||
Db 121492 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTYGGGTCCTGTTAAATGGATT 121553

Qy    804 ywsnttyathmgngargcngarcayaarwswsngaraaaaytcayccngayaayagtnat 863
          ::| |||||:|| |||:|.|| ||::||:|:|:|||||||:|||:|:||| ||:|:|:||
Db 121552 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGGAAAAATTTGCAGCCCTGAGAAGTGAT 121493

Qy    864 haaraaaraaraaayccnttywsngarggngaarttyaarytngcngcngaarathtgyathtg 923
          ||:|||:|||:||| |||||:.|||:||||:|||||:||:||:||:|| |||:||||| ||||
Db 121492 AGAAAACAGAAAAAAACCCATTTTTCTGAGGAGAAATTCAATCAGCCGCTGCAGAAAATTTGCATAAAG 121433

Qy    924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
          :||:|||||:||:||:|||||:||:|||||:||:|| |||:|:||:||:|||||:||:||
Db 121432 TAACAAGAAGCTGAATGTTAAYCCGCAAGACAATAGGGAAAATGTCTCCAGGGCATGTCA 121373

Qy    984 rmgnwsiwsnarcarwsnathaarwsnytngcntggngccnmgnmgnaarathtggttytg 1043
          |:|:|:|:|:|:|||:|:||:||:|:||:||||||:|||:|:|:|:|||:|||||||||||
Db 121372 GAAGTCTTCACAGCAGCCGCCCTCCCATCACAGGCCTGGAGCCTAGGAGAAAATGGTTTTG 121313

Qy   1044 yggnacnggngcncgggnwsnytntgytgygtncarccnmgnngayytngtnccntggtncc 1103
          :|||||:||:||:||||:|:|:|:||:||:||:||:||:|:||||::|:||:||:||||:||
Db 121312 TCGGCCACAGGCCCAGGGTCCCAAATGCTGTGTCCAGCCTAGGGACTTGGTCGCGTGCATCCC 121253
```

```
Qy      1104 ngtaaaywsngcngtngcnwsnsaarg-------------------------g 1130
Db    121252 AGCCCCTCTAGCCAATGGCTCAAACGGGCCCAAATGTAGATACAGCTCAAGCTGTGGCTTCAG 121193
Qy      1131 ngcnwsnccnaarccntggcarytnccnwsnggngarccgtnggncnaararws 1190
             .:||||||||||||||||| ||||||||||| | |||||||:|||:::
Db    121192 AGGGTGCCTAAGCCTTGGCAGTTTCCACATGGTGTGTTGAGTCTTTAGGTACACAGAAGTC 121133
Qy      1191 nmgnathgargtntgggarccncnccnaathngnttycaaraarathtaygnaayncntggat 1250
             .:|||:||||| |||||||| |||||||| ||||:|||||| || |||||||
Db    121132 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAANGCTTGGAT 121073
Qy      1251 gccnmgncaaraatttygcngtngngngtnggngnwsnwsntggmgnacnwsngcnmgngngtngt 1310
             .:| ||| | |:| .| |||||||||||||||||| |||||||||||||||||:
Db    121072 GCCCAGCCACAGAAGTTTGCTCACGGGGCAGGGATCTCATGGGAACCTCTCCTAGGGCACT 121013
Qy      1311 ncaraargnaaygtngngntgggarccncncnncaymgngtnccnwsnggngcncnwsnws 1370
             .:|||||:||||| ||||||||| |:|||||||||| ||| ::
Db    121012 GCAGAAGGGAAAATGTGGGGTCGAAGCCACCATATNGGAGTCCCTACTCGGGCACTTCTAG 120953
Qy      1371 nmngngcngtnmgnmngnwsncnccnccnwsnwsmgnmytncaraargngmgnwsnacngayws 1430
             .:| .|  :| .| |||||||||||||||||:||||||| |
Db    121372 TGGAGCTGTGAGAAGAGGGGCCCAATGACCTTCCAGAACCCCGTGAAAGCAGCCAG 120893
Qy      1431 nytncaarcaaygtnccngarzaarwsnacngayacncaartgycarccngtnaaargngcngg 1490
             .:|: |||||| .|:|||||:|||||||:||||||||| |||||:
Db    120892 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 120833
Qy      1491 natggarwsngtnccntayearacngtngtncgngarytnacnaaracngtnggnnathta 1550
             .:| |||||||:|||||:||||| ||||||| || ||:|||||||||||||| |:|
Db    120832 GAGGGAGGCTGTGCCCTACNAAAGCCACAGGGGTGGAGTTGCCCAAGACACATGGGAAACCCA 120773
Qy      1551 yytnytncaytgtycaygayytngaxygtnmgncayggngtnaaarmgngaycayttcgggnc 1610
             .:|:| .:|||:||||||||||||:||||||| ||||||||||:
Db    120772 CCTCCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGGATCATTTTGGAGG 120713
Qy      1611 nytnmgnttygaytgyccnacngggnttymgnacntayatgggnccnngt-nccnytntgyt 1669
             .:|.| :|||||||:|||:|||| ||||||| ||||::||| |:|||:|||:|
Db    120712 TTTAAAATTTGATTGCCTTCGGACTTGCATGGGCCCTGTAACCCCTTGTT 120653
Qy      1670 tyggncarttytytyccnttyggnacncngcngtnttyacncar 1710
             .:|||:|||||||||:|||||| ||||||||||||||||:
Db    120652 TTTGCCCAATTTCTCCCATTTGGAACAGCGTGTATTTACCCAG 120612
```

```
Qy       744 rgargarathwsnaarcarcatwsnathcaargargtnacncgggtnytnytnaaargcntt 803
             ::|:||:|||:|||::| ||:||||||:::  ||||| |||  | |||| |||   || ||
```

Query Match   28.2%; Score 488.8; DB 2; Length 163674;
Best Local Similarity   48.5%; Pred. No. 2.8e-108;
Matches   485; Conservative 202; Mismatches 280; Indels 34; Gaps 2;

```
RESULT 10
LOCUS       AC015727   175639 bp    DNA     linear   HTG 09-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-420A6 map 17, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC015727
```

```
Db 136131 GGAAGAAATTCTAAGCAGCAAAACATTCAAGGTGACTTGGCTCCTGTGTTAAATGGATT 136190
Qy    804 ywsnttyathmgnargcngarcacyagarwsnwsngaraaygytncaycccngayaaaygtnat 863
          ::|::  ||:||  :|||||
Db 136191 CAGTTTTATTGAGGGAAGCAGCGACATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 136250
Qy    864 haaraaraaraayccnttywsngargggnaarttyaaarytngcngcngarathtgyathtg 923
Db 136251 AGAAAAGAAAAAACCCATTTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAAATTTGCATAAG 136310
Qy    924 yaaygargaryatnaaygtnaayccncargayaaygyggnaraayathwsntggacntgyca 983
Db 136311 TAACAGAGACTGAAATGTTAATCCGCAAGACAATAGGGAAAAAGTCTCCAGGCCATGTCA 136370
Qy    984 rmgnwsnwsnwsncarcarwsnathaarwsnytngcngtngmgncncnmgnmgnaaatggtytg 1043
Db 136371 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCCTGAGGACTGGAGAAAAATGGTTTTG 136430
Qy   1044 yggnacnggnccnggnwsnytntgytgygtncarccnmngnayytngtncntgygtncc 1103
Db 136431 TGGGCCAGGCCCAGGGTCCCAACGCTGTGTCCAGCCTAGGGACTTGGTGCGCGTGCATCCC 136490
Qy   1104 ngtnaaygtnaaysngcngtngcnwsnsngarg------ 1130
                                              -g
Db 136491 AGCCGCTCTAGCCATGGCTGTGAAAGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 136550
Qy   1131 ngcnwsncnaaccccntggcaarytncncnwsnsngngtngarccngtngngcnaaraarws 1190
Db 136551 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTCAGTGCTTTTAGGTACACAGAAGTC 136610
Qy   1191 nmgnathgargtntggarcccncnathmgnttycaararaarathtaygngnaayccntggat 1250
Db 136611 AAGAAATTGAGGTTTGGGGAACTTCCACCCTGGATTTCAGAAGAGTGTATGGAAAACGCTTGGAT 136670
Qy   1251 gccnmgncncaaraarttygcngtngngngtnggnwsnswsntggmgnacnwsngcnmgngtngt 1310
Db 136671 GCCCAAGCCCAGAAGTTTTGGCTACGGGGATCTCATGGAGAAACCTCTPCGCTAGGGCACT 136730
Qy   1311 ncararaargnaaaygtngnttgggngntggagarccncncncaymgngtnccnwsngggngcncnwsnwws 1370
Db 136731 GCAGAAGGGAGAAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGCGCACTTTCTAG 136790
Qy   1371 nmgngncngtnmgnamgnwsncnccnwswsnmgnytncararaygngnmgnwsnacngayws 1430
Db 136791 TGGAGCTGTGAGAGAGGCCCAAATGACCTCCAGAATGGTAGATCCACCTACAG 136850
Qy   1431 nytncarcaygtncngaraarwsnacngayacncaartgycarccngtnaargcngcngg 1490
Db 136851 CTTGCACCGTATGCCTGGAAAAAGCTMGCAGACACTCAACACGCCGTGAAAAGCAGCCAG 136910
Qy   1491 natggarwsngtnccntayaaracngtngngwsnwsntggmgnacntayggnnathha 1550
Db 136911 GAGGGAGGCGTGCCCTACAAAGCCACAGGGTGGAGTTGCCCAAGACCATGGGAACCCA 136970
Qy   1551 yytnytncaytgycaygayytngngayytngcaytngmgncaytcggnngc 1610
Db 136971 CCTCTTGCCATCACAGCCATGGATCGTGAGACCTGGAGTCAAAGAGAATCATTTTGGAGG 137030
Qy   1611 nytnmgnttygaytgyccnacngngnttyngnacntayatggnccngngt-ncnytntgyt 1669
Db 137031 TTTAAAATTTGATTCCGTTCCTGGTTGTCATGGGCCCTGTAACCCCTTTGTT 137090
Qy   1670 tyggncaartyyttyycnttyggnacngnacngngtntytyacncar 1710
Db 137091 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 137131
```

```
FEATURES             Location/Qualifiers
source               1..175639
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-420A6"
                     /clone_lib="RPCI-11 Human Male BAC"
  misc_feature       1..13347
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
  misc_feature       14048..14259
                     /note="assembly_fragment"
  misc_feature       14360..15029
                     /note="assembly_fragment"
  misc_feature       15140..15832
                     /note="assembly_fragment"
  misc_feature       15933..16689
                     /note="assembly_fragment"
  misc_feature       16790..18092
                     /note="assembly_fragment"
  misc_feature       18193..19993
                     /note="assembly_fragment"
  misc_feature       20094..22205
                     /note="assembly_fragment"
*     19994  20093: gap of    100 bp in length
*     20094  22205: contig of 2112 bp in length
*     22206  22305: gap of    100 bp in length
*     22306  23915: contig of 1610 bp in length
*     23916  24015: gap of    100 bp in length
*     24016  26544: contig of 2529 bp in length
*     26545  26644: gap of    100 bp in length
*     26645  28851: contig of 2207 bp in length
*     28852  28951: gap of    100 bp in length
*     28952  30454: contig of 1503 bp in length
*     30455  30554: gap of    100 bp in length
*     30555  33015: contig of 2461 bp in length
*     33016  33115: gap of    100 bp in length
*     33116  35727: contig of 2612 bp in length
*     35728  35827: gap of    100 bp in length
*     35828  59255: contig of 23428 bp in length
*     59256  59355: gap of    100 bp in length
*     59356  62539: contig of 3184 bp in length
*     62540  62639: gap of    100 bp in length
*     62640  67930: contig of 5291 bp in length
*     67931  68030: gap of    100 bp in length
*     68031  73174: contig of 5144 bp in length
*     73175  73274: gap of    100 bp in length
*     73275  77427: contig of 4153 bp in length
*     77428  77527: gap of    100 bp in length
*     77528  81788: contig of 4261 bp in length
*     81789  81888: gap of    100 bp in length
*     81889  86379: contig of 4491 bp in length
*     86380  86479: gap of    100 bp in length
*     86480  91547: contig of 5068 bp in length
*     91548  91647: gap of    100 bp in length
*     91648  97783: contig of 6136 bp in length
*     97784  97883: gap of    100 bp in length
*     97884 105190: contig of 7307 bp in length
*    105191 105290: gap of    100 bp in length
*    105291 111855: contig of 6565 bp in length
*    111856 111955: gap of    100 bp in length
*    111956 119820: contig of 7865 bp in length
*    119821 119920: gap of    100 bp in length
*    119921 131299: contig of 11379 bp in length
*    131300 131399: gap of    100 bp in length
*    131400 139546: contig of 8147 bp in length
*    139547 139646: gap of    100 bp in length
*    139647 155538: contig of 15892 bp in length
*    155539 155638: gap of    100 bp in length
*    155639 173481: contig of 17843 bp in length
*    173482 173581: gap of    100 bp in length
*    173582 175639: contig of 2058 bp in length.
  misc_feature       22306..23915
                     /note="assembly_fragment"
  misc_feature       24016..26544
                     /note="assembly_fragment"
  misc_feature       26645..28851
                     /note="assembly_fragment"
  misc_feature       28952..30454
                     /note="assembly_fragment"
  misc_feature       30555..33015
                     /note="assembly_fragment"
  misc_feature       33116..35727
                     /note="assembly_fragment"
  misc_feature       35828..59255
                     /note="assembly_fragment"
  misc_feature       59356..62539
                     /note="assembly_fragment"
  misc_feature       62640..67930
                     /note="assembly_fragment"
  misc_feature       68031..73174
                     /note="assembly_fragment"
  misc_feature       73275..77427
                     /note="assembly_fragment"
  misc_feature       77528..81788
                     /note="assembly_fragment"
  misc_feature       81889..86379
                     /note="assembly_fragment"
  misc_feature       86480..91547
                     /note="assembly_fragment"
  misc_feature       91648..97783
                     /note="assembly_fragment"
  misc_feature       97884..105190
                     /note="assembly_fragment"
  misc_feature       105291..111855
                     /note="assembly_fragment"
  misc_feature       111956..119820
                     /note="assembly_fragment"
  misc_feature       119921..131299
                     /note="assembly_fragment"
  misc_feature       131400..139546
                     /note="assembly_fragment"
  misc_feature       139647..155538
                     /note="assembly_fragment"
  misc_feature       155639..173481
                     /note="assembly_fragment"
  misc_feature       173582..175639
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
```

```
BASE COUNT   48575 a   37483 c   38030 g   48544 t   3007 others

Query Match         28.2%;   Score 488.8;  DB 2;  Length 175639;
Best Local Similarity 48.5%;  Pred. No. 2.9e-108;
Matches 485;  Conservative 202;  Mismatches 280;  Indels  34;  Gaps  2;

Qy    744 rgargarathwsnaarcaarwsnathcargargtnacntgggtnytnytnaargcntt 803
          ::||||:||:||:::: ||:::|:|| ||:|||:|  || ||||| || || || ||
Db   2929 GGAAGAAAATTTCTAAGCACCAAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 2870

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          :: ||:||| :|||:|:|:|||:||||||||:||| :||:| ||:||:: |::||:|||
Db   2869 CAGTTTTATGAGGGAAGCAGCAGCATAAAAGTTCAGAAAAAITTGCAGCCTGAGAATGTGAT 2810

Qy    864 haataararaaycncnttywsngrcgrgnaarttyaaaryytngcncngaraathtgyathg 923
          :||||:|||:||:||:|| |:| |::|| ||:||||| ||:|:|:||::|||:||||||:||
Db   2809 AGAAAAAGAAAAACCCATTTTCTGAGGAGAAAATTCAATCAGGCTGCAGAAATTTGCATAAG 2750

Qy    924 yaaygaargarytnaaygtnaayccncarcayayayaayggngaraaayathwsntggacntgyca 983
          :|||:| |::::|:|| |::||::: |||:|| ||||: |||||: |::|| :::| | || ||
Db   2749 TAACAAGAAGACCTGAATGTTAATCCACAACGACAATGGGGAAAAATGTCTTCCAGGGCATGTCA 2690

Qy    984 rmgnwswsncaarcarwsnathaarwsnytngcntcggmgnccnmgnaartggttytg 1043
          :: :: :: || ||:|:||:: | :|::| :|:|:| :: ||:||:||||||:|
Db   ...
```

```
Db  2689 GAAGTCTTCACACAGACCCCTCCATCACAGGCTGGAGGCCCTAGGAGGAAAATGGTTTTG 2630
         |:||  ||| ||    ||: || |    |:| ||| || |  | ||   || | ||| |
Qy  1044 yggnacnggnccnggnwsnaytntgyttgygtncaarccnmngnayytngtnccntgyytncc 1103

Db  2629 TGGGCCAGGCCCAGGGGTCCCAAATGCTGTGTCCAGCCCTAGGGACTTGGTGCGCTGCCATCCC 2570
         ::|::|| ::|||: |::||  :|||||:||:| :||||::|| :| ||:|| || ||
Qy  1104 ngtnaaywsngcngtngcnwsngarg-------------------g 1130

Db  2569 AGCCGCGCTCTAGCCATGCCTGAAAGGGGCCCAATGTAGRATAGAGACTCAAGCTGTGGCTTCAG 2510
                 ||  ||  |  | ||   |      |:
Qy  1131 ngcnwsncnaarccntggcarytnccnwsnggngtnggarccngtnggngcnaaraarws 1190

Db  2509 AGGGTGCCCTAAGCCCTTGGCAGTTTCCACATGGTTGTTGAGTCTTTAGGTTACACAGAAGTC 2450
         |:::  || |  | :| |||  || ::|  ||| ||| ||  | :| | || ||  |:|:||
Qy  1191 nmgnathgargtntggarccncacnathmgnttycaraaaratntaygngnaayccntggat 1250

Db  2449 AAGAATTGAGGTTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATNGAAACCGCTTGGAT 2390
         ||| |  ||| |||| |||  |||  |   |:| |:||||| |:|  | |||  ||||||| |
Qy  1251 gccnmgncararttygcngtngngnwtnggngwsnwsntggmgnacnwsngcnmgngmgngtngt 1310

Db  2389 GCCCAGCCAGAGTTTCCTGCAGGGGCAGGGGATCTCANGGAGAACCCTCGTCGTAGGGCACT 2330
         :|| | ||| |: :||| :||||  | |||: ||| |:| ||| ||  | :|  | |||
Qy  1311 ncaraarggnaaygtnggntnggarccncncaymgngtnccnwsnggngcncncnwsnws 1370

Db  2329 GCAGAAGGGAAAATGTGGGGTGCGNAGGCCGTCGNAAGCNCCACCATARGGAGTCCCTACTCGGGCNCACCTTCTAG 2270
         |||   || |  ||| | | |  |  ||:|| | :|| |  || ||  |  || ||  || | || |:|  || |
Qy  1371 nmgngngcngtnmgngmgnvsnccncncnwsnwsmgnytncaraargngmgnwsnacngayws 1430

Db  2269 TGGAGCTGTGAGAAGAGGGCCCAATGACCTCCAGAACCTGGTAGNATCCACCTACAG 2210
         :|| ||| |  ||| |||  |  | ||| :|| | ||   ||| ||:|  ||| |
Qy  1431 nytncarcacaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1490

Db  2209 CTTGCACACGTATGCCTGAAAAACCTGCAGACACTCACACCCCGTGAAAAGCAGCCAG 2150
         |||| || |||| ||||| |: |||||| ||| |  | ||  |||| | ||||| |
Qy  1491 natggaarwsngtncntayaaaracngtngtncngarytnacnacaacngtnggnathta 1550

Db  2149 GAGGGAGGGCTGTGCCCTACAAAAGCCACACAGGGTGGAGTTGCCCAAGACCATGGGAACCCA 2090
         | ||| || ||| |  ||||| || || ||| || |  || ||| || || | || | |||
Qy  1551 yytnytncaytgycaygayytngaygtnaarmgngaaycaytgyggngc 1610

Db  2089 CCTCTTGCATCAGCATGACCTGGATGTGAGACCGTCAAAAGGATCATTTTGGAGG 2030
         ||| |||| |  |  || ||| | :| | |||:|| ||  | | ||| ||| ||
Qy  1611 nytnmgnttygaytgyyccnacngnttymgnacntayatsggnccngt-nccnytntgyt 1669

Db  2029 TTTAAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCANGGGCCCTGTRAACCCTTTGTT 1970
         |||||||  ||||| |||| |  ||  | ||  |||:| |||||||| || ||| |||:|:
Qy  1670 tyggncaarttytyccnttyggnacngncngngtnttyacncar 1710

Db  1969 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 1929
         |||||||||||:|||:|| || ||||| |||  | || ||
Qy  1711 ...
```

```
RESULT 12
AL353634
LOCUS       AL353634              77405 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome X clone RP11-435A2 map q21.33-22.3, ***
            SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AL353634
VERSION     AL353634.4  GI:9863649
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 77405)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9213111.
            ---------
            Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            Project Information
            Center project name: bA435A2
            ----------
            Summary Statistics
            Sequencing vector: XGAP4; version 4.5
            Assembly program: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 13% of reads Chemistry:
            Dye-terminator Big Dye; 86% of reads
            Consensus quality: 76604 bases at least Q40
            Consensus quality: 77098 bases at least Q30
            Consensus quality: 77320 bases at least Q20
            Insert size: 77405; sum-of-contigs
            Insert size: 85298; 1.5% error; agarose-fp
            Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
            coverage: 4.30x in Q20 bases; agarose-fp
            ----------
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..77405
                     /organism="Homo sapiens"
```

```
Query Match   28.2%;  Score 488.6; DB 2;  Length 181842;
Best Local Similarity 47.5%;  Pred. No. 3.2e-108;
Matches  486; Conservative 212; Mismatches 290; Indels  36; Gaps  2;

Qy    743 argaargarathwsnaaarcarcarwsnathcaargargtnacntgggtnytngytnaaargcnt 802
          :|| ::| || ||:|| |::|||| :| ||:|| | ||:|::|| |||| :| :|| | || |
Db  64835 AAGAAGGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTTGTTTAAAGGCAT 64776

Qy    803 tywsnttyathmngnargcngnarcayaarwsnwsngaraayytncayccngayaaygtna 862
          || ||::|| ||:||| | ||| :|| |:| ||| |||||::||:|| |||::|||:| :|
Db  64655 GTACCAAGGAGCCTAATGTTAATCCCAAGACCATCGGGAAATGTCTACAGGCCATGTC 64596

Qy    863 thaararaaraaycccnttywsngaarggnaaarttyaarytngcngcngarathgyatht 922
          :|| | :| :||||:| | ||||:::| |||||::||:||| |||:|||| | ::|||:|
Db  64715 TAGAACAAGAAAAACCCATTTTCTGGGGAGAAAATTCAAGCCTATTGCAGAAAATTGCCTAA 64656

Qy    923 gyaaygargaryytnaaycccnccargayaaygggaraaayathwsntgacntgyc 982
          | ||| |||:|| ||:|| :|||:|| ||:| ||| | ||| || :| || ||:|
Db  64655 GTACCAAGGAGCCTAATGTTAATCCCAAGACCATCGGGAAATGTCTACAGGCCATGTC 64596

Qy    983 armgnwswsncarcarwsnathaarwsnytngcntggmgnccnmgnmng---naartggt 1039
          |:|| ||| || :| |||||:| ||| |:|::| ::|:||| | :|| |||| ||:
Db  64595 AGGGAACTTCATGGCTGCCCCTCCCATCACAACCCTGCAGGCCCCAGGAGTAAAAGTGGT 64536

Qy   1040 tytytgyggnacngggncnggnwsnytntgytgygtncarccnmgnagnayytngtnccntgyg 1099
          |:||| |:| ||::|| :| || |:| :| ::| :| ||:|| |||| :|:| | || ::|
Db  64535 TTCATGGCCAGGGCCAAAGGGCCGTTCCTGCGTGTGCAGCACAGGGACTTGGTGCCCTGTG 64476

Qy   1100 tnccng----------tnaaywsncgngtngcnwsng 1126
          | ||| :|:||:|||:|:| ||||| |
Db  64535 TTCATGGCCAGGGCCAAAGGGCCGTTCCTGCGTGTGCAGCACAGGGACTTGGTGCCCTGTG 64476

Qy   1127 arggngcnwsncnaarccnntggcaryntnccnwsnggngtngarccngtngggngcnaara 1186
          :|||:|||:|| |:|| :| |||| |:|:|| ||||:|||| :|| ||||| ||||: ||:
Db  64415 AGGGTGCAAGACCCAAGCCTTGGCAGCTTCCACGTGATGTTGAGCCTGTGGGTGCACAGA 64356

Qy   1187 arwsnmgnathgargtntgggarccncnaathmgnttycaraaararathtaygggnaayccnt 1246
          |: || |||:|:| :||| |:||| :| || :|||:| ||||:|||| ||||||::||::|
Db  64355 AGTCAAGAACTGAGGTTGGGAACCTCTCCTAGAATTTCAGAAGACAGTGTATGGAAACGCCT 64296

Qy   1247 ggatggccnmgncaraaarttygcngtngggntngngnwswsntggmgnacnwsngcnmgng 1306
          |||||:|| ||:| |||:||||:|||| :||:|||:| || ||:|| :| :|| ::|:||
Db  64295 GGATGCCCAGGCAAAAGTTTGCTCGGGTGTGGGACCCTCATGGAGAACCTCTCGTAGG 64236

Qy   1307 tngtncaraarggnaaygtnggntggggarccncncncayrmgngtncccnwsnggngnccncnw 1366
          || |:||||:|| |||| :||:|||||::| :| :|:||:||:| | |||||:|||:|| | :|
Db  64235 CAGTGCAGAAGGAAAAATGTGGGGTCAGAAGCCCACCAGAGTCCCTACTGCACCAC 64176

Qy   1367 snwsnmgncgntnmgnmgnwsncnccnwsnsnmgnytncaraarggngmgnwsnacng 1426
          :|:|: || ::| ||:| :|| || || :|:|: |::| | |: ||:|:||:|: |||
Db  64175 CTAGTGGGAGCTGTGAGAAGGGGGGCCACTGTCCTCAGACCCCAGAGATGATAGATCCACTG 64116

Qy   1427 aywsnytncarcaygtnccngararatwsnacngaycancartgycarccngtnaargcng 1486
          |:|: :|| || ||:||| :|||||:|| || ||:| ||| |:|| ||| ||:|||||||:
Db  64115 ACAGCTTGCACCATTCACCTGGAAAAAGCCACAGAACACTCAACCGCCATGAAAGCAG 64056

Qy   1487 cnggnatggarwsngtnccntayaaracngtngtgcngcngarytnacnaacngtnggna 1546
          |:|||||||:||:|||||:|| |||||| :|||:| ||| || :|:| || ||:|||
Db  64055 CCAGGAGGGAGGCGTGTACCCTGCCAAAGCCACAGGGGTGGAACTGCCCAAGACTATGGGAA 63996

Qy   1547 thtayytnytncaytgycaygayytngaytnmncayggngtnaarmgmgnayccayttyg 1606
          :|| :|:|: |||| ||:|||::|||||:| :| || |:||| |:|:||| |:|||||| |
Db  63995 CCTACCTCTTGCATCAGCATGACCATGACCTGGATGTGGTGTCAAAGGAGAGATCATTTTG 63936

Qy   1607 gngcnytnmgnttygaytgycnacnggngttymgnacntaygtggnccnngtnccncnytnt 1666
          |:||:|||:|||:|| |:|:|| | :|||:|:|| || |:|||| ||:| || || :|:|
Db  63935 AAGCTTTAAAATTGACTGCCTCACTGGATTTCAGACTTGCCATGGGCCCTGTACCACTTT 63876

Qy   1667 gyttyggncarttytyccnttyggnacngcngngtnttyacncaartgyytnayytncayt 1726
          |:||:||:| |::|:|:| ||:||:|| :|:| | |::|| ||| |:|:: |:|:|||||
Db  63875 GTTTTGGCCAARTTCTCCCAATTTGGGAGGGCTATATTTACCCCAGTACTTGTACCCCATT 63816

Qy   1727 gyat 1730
          :|||
Db  63815 GTAT 63812
```

```
ORIGIN
BASE COUNT     53809 a  39941 c  37882 g  49105 t  1105 others

     misc_feature   173204..181842
                    /note="assembly_fragment:01461"

     misc_feature   173103..173203
                    /note="assembly_fragment:00742"

     misc_feature   155711..173103
                    /note="assembly_fragment:00583"

     misc_feature   155610..155711
                    /note="assembly_fragment:01445"

     misc_feature   126687..155610
                    /note="assembly_fragment:01651"

     misc_feature   107745..126586
                    /note="assembly_fragment:00035"

     misc_feature   107644..107745
                    /note="assembly_fragment:01169"

     misc_feature   105511..107644
                    /note="assembly_chain:1"

     misc_feature   105410..105511
                    /note="assembly_chain:1"

     misc_feature   94703..105410
                    /note="assembly_chain:1"

                    /note="assembly_chain:1"
```

```
BASE COUNT    25072 a   14551 c   15133 g   22649 t
ORIGIN
misc_feature  1..77405
              /note="assembly_fragment:00554"
              /clone_lib="RPCI-11.2"
              /clone="RP11-435A2"
              /map="q21.33-22.3"
              /chromosome="X"
              /db_xref="taxon:9606"
```

```
Query Match   28.2%;  Score 488.2;  DB 2;  Length 77405;
Best Local Similarity   47.0%;  Pred. No. 3e-108;
Matches 508;  Conservative 225;  Mismatches 310;  Indels 37;  Gaps 4;

Qy    688 ytncaycaytgyaargtnaayathtggytnatgmgnaarcararthytngcnaayaargar 747
          |  |||  :| ::|| || ||| ||  |: | | |:|  | |  | ||| :|    ||
Db   8637 CTGCCCTAGAGATCTGTGGAATTTGACCTTGAGAGAGATGAATGCATTACAGGTGGAA 8696

Qy    748 garathwsnaarcarcaarwsnathcargargtnacntgggtnaytnytnaargcnttywsn 807
          || |||:|: ||::|||| | |||::| || |||| || | |||: | |:|||:|  |||
Db   8697 GAAATTTCRAGCAGCAAAACCAATTCAAAGACGTGACTTGTGTCCCTGTTAAAGGCACTCAGT 8756

Qy    808 ttyathmgngaargcngarcayayarwsnswsngaraayytncayccngayaaygtnathaar 867
          || |:|||||||| ::: | || ||| ||| | :|| | :||||: :| | || |||
Db   8757 TTTGCAAGGGAAGCAGAACATAAAAGTTCACAAAAATTTTCAGCCTCACAATGTAGTAGAA 8816

Qy    868 aaraaraaycc-nttywsngarggnaarttyaarytngcngcnaarathtgyathtgyaa 926
          | | || |:| :|| | |:| ||||:|:||| | || ||   |:| : || | ||| |:|
Db   8817 AAGAAAAACCCATTTTTCTGATGAATAAATTGCGCAGAAATTGCATAAGTAA 8876

Qy    927 ygargarytnaaygtnaaycncacgayaaygnngaraayathwsntggacntgycaarmg 986
          | :|| || |:| :|| ||| ||| | || |: | |:|||  :: ||| ||| ||:|
Db   8877 CGAGGAGCCAAAATGTTAAATCACTAAGACACATGGGGAAAAAATATCTCCAATGGCTTGTGTCAGAG 8936

Qy    987 nwsnwsncacarwsnathaaarwsnytng--cntggmgnccnmgnmgnaartgttytgty 1044
          |: :: :||| ||| || | | ||   |: |:||   |  |  | :|| ||| ||:|:|
Db   8937 GTCTTCACGGCGAGCCCCTCACATCACTGGCCCCCAGAGGCCTAGGAGAAATGGTTTTAT 8996

Qy   1045 ggnacngnccnggnwsnytntgytgygtncarccnmgngayytngtcnccntgytgtnccn 1104
          || |||  ||: ::|||| |  || |: |:|| ||| || |||| | :||: | | |:||
Db   8997 GGGTCGGGCCCCAGGGTCCCCATGCTATGTGCAGTCTAGGAGACTTGGTCGTCGTCCCA 9056

Qy   1105 gtnaa------- ywsngcngtngcnwsngarggn 1131
          || || :: :|| |  | |  ::|:| |
Db   9057 GCCACTCCACCCATAACTGAAAGGGGGGCCAAGGTATAGCTTGGGCTCTTGCTTCAAAGGAT 9116

Qy   1132 gcnwsnwsncnaarccntggcaryytnccnwsngngyntngarccngtnggngnaararwsn 1191
          ||: | ||:|:||||| |  || || | |  |:| || :||: :|  | | |::|
Db   9117 GGAAGCCCCAGGCCTTGGCACGTTCCATGTGGTGTTGAGCCTGCACTGAAGTCA 9176

Qy   1192 mgnathgargtgtntggaarccncncnathmgnttycaraaarathtayggnaaycccntggatg 1251
          ||:||||  ::| |  ||  |     | ||  |||| |::|  | :|| :|| |: ||||
Db   9177 AGAATTGAGGTTTGGGAACCTCCACTCAGTAGAAGAGTATGGAAATGCCTGGATG 9236

Qy   1252 ccnmgncararttygcngtnggngtnggngnwsnwsntggngnacnwsngcnmgngtngtn 1311
          ||||| || ||| ||  ||| ||| | | || | ||:: || ||:|| || :| | |||:
Db   9237 CCCAGGCAAAAGTTTGCTGCAGGGGCGGAGAACCTCATGGGAGAACACTCATGGGGGGTG 9296

Qy   1312 caraarggnaaygtnggngntggaarccnwsnggngngcnccnwsnwsn 1371
          || ||  || :||| || |||  | |||  | || || | |||  | ||
Db   9297 TGGAAGGGAAAAGTGGGGTCATAGCCGCCACACAGAGTCCCTACCAGGGACCACCTAGT 9356

Qy   1372 mgngncngtngnmgnwsncncncnwsnswsnmgnytncaraarggngmgnwsnacngaywsn 1431
          ||: ||:||  |: |   || |   |  | ||:||||  ||| ||| || | |   ||:|
Db   9357 GGAGCTGTGAGAGAGGGGCCACCACCATCTCCAGACGAACCGGACAGC 9416

Qy   1432 ytncaraaygtnccngaraatrwsnacngayacncaartgycarccngtnaarcgncnggn 1491
          :|  ||:|| || || ||| | || ||| | || || |||  |:  ||| || ||  || |:
Db   9417 CTGCAACCGTGCTGCCTCGAAAAGCCCCGAGACACTCAATGCCAGGCCTGTGAAAAGCAGCAGG 9476

Qy   1492 atggarwsngtnccntayaaracngtngtngcngaryt  1551
          ||||:    || || ::| |  | |||  | || || |:
Db   9477 AGGGAGGCGTACCCTGCCAAAGCCACAGGGGGCGGAGAGCTGCCCAAGATCATGGGGACCAAT 9536
```

```
Qy   1552 ytnaytncaytgycaygayytngaaygtnngncaayggngtnaarmgnacaycattyggngcn 1611
          :| |:||| :|||: | || | ||  | ||| |||   || || ||||  ||  | |||  ||
Db   9537 CTCTTGCATCAGCATGCAGACCTGGATATGAGACATGGAGAGTCAAAGAGAAATCATTTTGGAGCT 9596

Qy   1612 ytnmgnttygaytgycacacnggngnttymgnacntayatgtggngnccngt-nccnytntgytt 1670
          :| |  ||| || |  |: | :|  || || ||: |  |:||: | | |:|:| ||:: ||
Db   9597 TTAAGATTTGACTGTCCCACTGGATCTCAGACTTCACAAGTGGGTGGGGCTTGTTGTT 9656

Qy   1671 yggncaarttgytyccnttyggnacngcngtnttyacnncaartgyytntayytnaaytgyat 1730
          :|| || |:|:|::| ::| || ||:||  |:|:|  |:|||| ||| |||| || |:|
Db   9657 TGGCCAATAATCTCCCATTTGGAATGGCTGTGTATTTATCCAAATGCCTGTACCCCCATTGTAT 9716
```

```
RESULT 13
HSJ633H17/c            124531 bp     DNA     linear     PRI 07-FEB-2000
LOCUS       HSJ633H17
DEFINITION  Human DNA sequence from clone 633H17 on chromosome 1p31.2-32.2.
            Contains a pseudogene similar to part of MTCO1 (Cytochrome C
            oxidase 1), MTCO2 (cytochrome c oxidase II), MTATP8 (ATP synthase
            8) and GOT2 glutamic-oxaloacetic transaminase 2, mitochondrial
            (aspartate aminotransferase 2) pseudogenes, ESTs and GSSs, complete
            sequence.
ACCESSION   AL049710
VERSION     AL049710.18  GI:5650682
KEYWORDS    HTG; aspartate aminotransferase; ATP synthase; Cytochrome C
            oxidase; glutamic-oxaloacetic transaminase; GOT2; MTATP8; MTCO1;
            MTCO2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124531)
  AUTHORS   Moore,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 29, 1999 this sequence version replaced gi:5566565.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 633H17. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            633H17 is from the library RPCI-4 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/. VECTOR: PCYPAC2 The
            following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.
FEATURES             Location/Qualifiers
     source          1..124531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p1.2-32.2."
                     /clone_lib="RPCI-4"
                     /clone="633H17"
```

```
repeat_region   1. .234             /note="AluSx repeat: matches 59. .292 of consensus"
misc_feature    complement(235. .546)  /note="match: GSS: Em:AQ055811"
repeat_region   559. .760           /note="match: GSS: Em:AQ055811"
repeat_region   1158. .2018         /note="MIR repeat: matches 74. .262 of consensus"
repeat_region   2026. .2590         /note="L1MA4 repeat: matches 5282. .6145 of consensus"
repeat_region   2879. .9441         /note="HERVL repeat: matches 4676. .5241 of consensus"
repeat_region   10000. .10124       /note="L1PA2 repeat: matches 1. .6146 of consensus"
repeat_region   10623. .10689       /note="MIR repeat: matches 34. .154 of consensus"
repeat_region   10690. .11014       /note="MIR repeat: matches 36. .102 of consensus"
repeat_region   11136. .11262       /note="L2 repeat: matches 2389. .2710 of consensus"
repeat_region   13420. .13522       /note="L2 repeat: matches 2349. .2473 of consensus"
repeat_region   13602. .13725       /note="MIR repeat: matches 114. .219 of consensus"
repeat_region   13739. .13987       /note="MIR repeat: matches 130. .262 of consensus"
repeat_region   14368. .14486       /note="L2 repeat: matches 2397. .2666 of consensus"
repeat_region   14656. .14764       /note="L2 repeat: matches 2585. .2705 of consensus"
misc_feature    14765. .14965       /note="MIR repeat: matches 124. .252 of consensus"
misc_feature    14966. .15150       /note="match: GSS: Em:B30406"
repeat_region   15157. .15818       /note="MIR repeat: matches 13. .217 of consensus"
repeat_region   16403. .16945       /note="MER39b repeat: matches 14. .579 of consensus"
misc_feature    16410. .16693       /note="match: GSS: Em:AQ454645"
misc_feature    16410. .17037       /note="match: GSS: Em:AQ357205"
misc_feature    complement(16459. .16861)  /note="match: GSS: Em:AQ309712"
misc_feature    16531. .16576       /note="match: GSS: Em:AQ197233"
repeat_region   18053. .18154       /note="23 copies 2 mer tt 74% conserved"
repeat_region   18377. .18492       /note="MIR repeat: matches 7. .107 of consensus"
repeat_region   18525. .18954       /note="MER39b repeat: matches 476. .574 of consensus"
misc_feature    19026. .19126       /note="match: GSS: Em:AQ210513"
repeat_region   19866. .20007       /note="MIR repeat: matches 10. .114 of consensus"
repeat_region   20273. .20334       /note="MIR repeat: matches 28. .169 of consensus"
repeat_region   20403. .20505       /note="LTR16C repeat: matches 90. .152 of consensus"
repeat_region   20508. .20690       /note="MER81 repeat: matches 1. .110 of consensus"
repeat_region   20692. .20725       /note="LTR16C repeat: matches 190. .376 of consensus"
repeat_region   20828. .20863       /note="L2 repeat: matches 2660. .2693 of consensus"
repeat_region   21172. .21304       /note="18 copies 2 mer ac 94% conserved"
repeat_region   21201. .21331       /note="MER5A repeat: matches 40. .167 of consensus"
repeat_region   21466. .21528       /note="MER5A repeat: matches 56. .188 of consensus"
repeat_region   21952. .22077       /note="MER54A repeat: matches 214. .285 of consensus"

repeat_region   22083. .22259       /note="L2 repeat: matches 2578. .2707 of consensus"
repeat_region   22280. .22393       /note="FRAM repeat: matches -2. .175 of consensus"
repeat_region   22394. .22748       /note="L2 repeat: matches 2541. .2645 of consensus"
repeat_region   22749. .22846       /note="MLT1B repeat: matches 1. .390 of consensus"
repeat_region   22864. .23010       /note="L2 repeat: matches 2645. .2745 of consensus"
repeat_region   23011. .23310       /note="L2 repeat: matches 2329. .2429 of consensus"
repeat_region   23311. .23391       /note="AluJb repeat: matches 1. .299 of consensus"
repeat_region   23433. .23730       /note="L2 repeat: matches 2429. .2521 of consensus"
repeat_region   23731. .25397       /note="MSTD repeat: matches 66. .394 of consensus"
repeat_region   25477. .25556       /note="L1PB1 repeat: matches 4442. .6123 of consensus"
repeat_region   25557. .25651       /note="MSTD repeat: matches 1. .85 of consensus"
repeat_region   25745. .25785       /note="L2 repeat: matches 2585. .2692 of consensus"
repeat_region   26698. .27373       /note="MER58A repeat: matches 2702. .2742 of consensus"
repeat_region   27378. .27483       /note="MER31A repeat: matches 1. .661 of consensus"
repeat_region   27484. .27868       /note="MER93 repeat: matches 292. .397 of consensus"
misc_feature    27537. .27688       /note="match: GSS: Em:AQ127976"
repeat_region   27912. .28051       /note="MER93 repeat: matches 1. .188 of consensus"
repeat_region   29900. .30121       /note="L1PA13 repeat: matches 6011. .6152 of consensus"
repeat_region   30122. .30487       /note="HERVL repeat: matches 3020. .3235 of consensus"
repeat_region   30488. .30578       /note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region   30579. .30796       /note="HERVL repeat: matches 2933. .3020 of consensus"
repeat_region   30797. .31887       /note="MER58A repeat: matches 1. .221 of consensus"
misc_feature    complement(31888. .32090)  /note="HERVL repeat: matches 1792. .2933 of consensus"
misc_feature    32922. .33453       match: GSS: Em:AQ759038; match: GSS: Em:B58153; match: GSS: Em:B66544;
repeat_region   33175. .33412       /note="match: GSS: Em:AQ751315"
repeat_region   33454. .33885       /note="L2 repeat: matches 2473. .2708 of consensus"
repeat_region   33896. .33946       /note="MLT1C repeat: matches 6. .466 of consensus"
repeat_region   33947. .34302       /note="MIR repeat: matches 162. .215 of consensus"
repeat_region   34303. .34334       /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   34335. .34703       /note="MIR repeat: matches 129. .162 of consensus"
repeat_region   34704. .34830       /note="THE1C repeat: matches 1. .369 of consensus"
repeat_region   34831. .35552       /note="MIR repeat: matches 20. .129 of consensus"
repeat_region   36612. .36657       /note="L1MC2 repeat: matches 5573. .6325 of consensus"
repeat_region   36935. .37004       /note="L2 repeat: matches 2675. .2723 of consensus"
repeat_region   37383. .37648       /note="L2 repeat: matches 2638. .2707 of consensus"
```

Query Match   28.2%; Score 488.2; DB 9; Length 124531;

```
Db   95072 GTAT 95069

RESULT 14
LOCUS       HSU212C1                40714 bp    DNA         linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from cosmid U212C1, between markers DXS366 and
            DXS87 on chromosome X.
ACCESSION   Z69722
VERSION     Z69722.1  GI:1204119
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40714)
  AUTHORS   Odell,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquiries: humpub@sanger.ac.uk
COMMENT     IMPORTANT: This sequence is the entire insert of clone U212C1. The
            true left end of clone U212C1 is at 1 in this sequence. The
            right end of clone U212C1 is at 40714.
            U212C1 is from the human chromosome X-specific cosmid library.
FEATURES             Location/Qualifiers
     source          1..40714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="LL0XNC01-212C1"
                     /clone_lib="LL0XNC01"
     repeat_region   10..233
                     /note="L1 element fragment"
     repeat_region   388..758
                     /note="L1 element fragment"
     repeat_region   932..1108
                     /note="L1 element fragment"
     repeat_region   2571..2617
                     /note="L1 element fragment"
     repeat_region   2613..3503
                     /note="MER43 element fragment"
     repeat_region   5204..5229
                     /note="L1 element fragment"
     repeat_region   5230..5427
                     /note="MER43 element fragment"
     repeat_region   8954..9088
                     /note="13 copies of 2 mer 100 % conserved"
     repeat_region   12171..13985
                     /note="L1 element fragment"
     repeat_region   14179..14612
                     /note="L1 element fragment"
     repeat_region   14662..14780
                     /note="L1 element fragment"
     repeat_region   14824..15044
                     /note="L1 element fragment"
     repeat_region   15234..15369
                     /note="L1 element fragment"
     repeat_region   15421..15849
                     /note="L1 element fragment"
     repeat_region   16026..16155
                     /note="L1 element fragment"
     repeat_region   16073..16149
                     /note="THR1B element fragment"
     repeat_region   16204..16495
                     /note="THR1B element fragment"
     repeat_region   16532..16609
                     /note="Alu repeat: matches 1..308 of consensus"
     repeat_region   16624..16722
                     /note="THR1B element fragment"
     repeat_region   16626..16737
                     /note="MSTC element fragment"
     repeat_region   16651..16712
                     /note="MSTA element fragment"
```

Best Local Similarity 47.4%; Conservative 213; Mismatches 290; Indels 36; Gaps 2;
Matches 485; Pred. No. 3.5e-108;

```
Qy    743 argargaratwsnaarcarwsnathcargargtnacntgggtnytnaargcnt 802
          |:||||||||||||||||: ||||||||||||:|:|| :||::||||
Db  96092 AAGAAGGATTNCCTNAAGCACNAAAGCATTTGAGATNGACTTTGGTGTTAAAGGCAT 96033

Qy    803 tywsnttyathmgngargcngaccayaatwsnwsngataaaytyncayccngayaaygtna 862
          |::: |||| ||:|||||| |||||||||||||||||||:| ||||||:|:|| |
Db  96032 TCAGTTTTATAAGGGAAGCAGAGCATAAAAATTTGGAAAAATTTGCAGCCTGACTATGCAA 95973

Qy    863 thaaraaraaraayccnttywsngargnaarttyaarytyncngararthtgyatht 922
          |||||||:||||||||||| ||||||||||:||||||||:||||
Db  95972 TAGAAAAGAAAAAACCCATTTTCTGGGGAGAAACCTATTGCAGAAATTTGCCTAA 95913

Qy    923 gyaaygargaryytnaaygtnaayccncargayaayggngaaraayathwsntgaacntgyc 982
          ::: |||||||||||||||:||| ||||: ||||||||:||||||||||||||
Db  95912 GTAGCAAGGAGCCTAATGTTAATCCCCAAGACANCATCGGGGAAATGTCTACAGCCCATGTC 95853

Qy    983 armgnwsnwsnacarwsnathaarwsnytngcntggmgncncnmgng---naarttgt 1039
          |::: |||||||||:|||::||||||||||||||| || |||:|||
Db  95852 AGGGAAACTTCATGGCTGCCCTCCCATACACAAGCCTGCAGGCCCCCAGGAGTAAAAAGTGGT 95793

Qy   1040 tytgyghgnacngcngguwsnytntgtygyghgtncarccnmgngayytngtncntgyg 1099
          ||: :||||:||||||||||||:||:|||:||||||||||:||||||||||||||
Db  95792 TTCATGGGCCCAGGCACCCAGGGTTCCTGTGTCGTGTGGGGACACAGGGACTTGGTGCCCTGTG 95733

Qy   1100 tnccng----------------tnaaywsncngtngcnwsng 1126
          | ||| ||||| |||| ||: ||||||
Db  95732 TCCCAGCTGCTTCCAGCTATGGGCTGAAAGGGGCCAAGGTAGGAGGTGGGCTGGCTTCAG 95673

Qy   1127 argngncnwsnccnaarccnttggaarytnccnwsngngngtngarccngtnggghgncnaara 1186
          ::: |||: ||||||| |||||||||:|||||||| ||||||||:|
Db  95612 AGGGTGCAAGACCCCAAAGCCTTGGCAGCTTCCACGTGCGATCGTTGGGGTCACAGA 95613

Qy   1187 arwsnmgnathgargtntgggarccnconathmgnttycararararthtaygngnaaycnt 1246
          |:||||||||||||||||||| ||:|:||||||||||||||||:||
Db  95612 AGTCAAGAACTGAGGTTTGGGAACCTCTCCCTAGANTTTCAGAAGANGTAATGTAAAAAGTGGT 95553

Qy   1247 ggatgccnmgncarraartygcngtngguwsnywsntggmgnacnwsngcnmgng 1306
          |:||| |||:||||||||||||||| ||:|||||||:||||||||||||||||
Db  95552 GGAHGCCCAGGCAAAAAGTTTGCTTGCCAGGTGTGTGGGGACCCTCATGCTAGGG 95493

Qy   1307 tngtncaaraargnnaaygtnggntngggarccnccncacaymgntgtnccnwsnggngcnccnw 1366
          ::: |||||||||||:: ||:| |||:|||| |||||||||||| ||::
Db  95492 CAGTGCAGGAAGGAAAATGTGGGGTCAGAGCGCCCTACTGGGGCACCAC 95433

Qy   1367 snwsnmgngcngtnmgnmngnwsncccncnwsnsnmgnytncaraargngmgnwsnacng 1426
          ::: |||||||||||||||||||:| |||| ||| |:||:|:||:|
Db  95432 CTAGTGGACGTGTGAGAAGGGGGCCACTGTCCTCCCAGCAGNATGATGAGATCCACTG 95373

Qy   1427 aywsnytncarcaytnccngararwsnacncngaryacncaartgycarccngtnaargcng 1486
          |:||||||||||||||||:|:||||:|||:|||||:||||||:|||
Db  95372 ACACGTTGCACCATTCACCTGGAAAAGGCACACACNGGAAAGCCCCATGAAAAGCAG 95313

Qy   1487 cnggnatggaarwsngtnccnatayaaracngtngtcngngaryacnaaracngtnggna 1546
          |:||||||||||||||||||||||||:||||||||:|:|||||||
Db  95312 CCAGGAGGCAGGCGTGTACCCTGCAAAGCCACAGGGTGTGGAAACTGCCAAGAGTATGGGAA 95253

Qy   1547 thtayytnytnaytcnaytgtgcaygayytngagyntgnacyttmgnacntaacngtncccnytnt 1606
          |:||||||||||||||||::|:|||||||||:||||||| |||| |
Db  95252 CCTACCTCTTGCACAGCATGAGTTCGGTGTGTCAAAAGGGTGAAAGACTATTG 95193

Qy   1607 gngcnytnmgnttygaytgycncacngngnttymgnacntayatggncncngtnccnnytnt 1666
          |:||| ||||||| ||||||:: |||||||:||||||||||| ||:|||||||
Db  95192 AAGCTTAAAAATTTGACTGCTCCACTGGATTCAGACTTGCATGGGCCCTGTACCACACTTT 95133

Qy   1667 gyttyggncarttyttyccnttygnacncgngtnttyacncaartgyytntatayytncayt 1726
          |:||||||||||||| |||||||||||||||||||||::|
Db  95132 GTTTTGGCCAATTTCTCCCAATTTGGGAGGGGCTATATTTACCCAATACTTGTTACCCCATT 95073

Qy   1727 gyat 1730
          |:||
Db  95072 GTAT 95069
```

```
repeat_region    /note="THE1B element fragment"
                 17060..17653
repeat_region    /note="L1 element fragment"
                 19720..19797
repeat_region    /note="MER21B element fragment"
                 19877..19915
repeat_region    /note="MER21B element fragment"
                 20007..20256
repeat_region    /note="MER21B element fragment"
                 20518..20734
repeat_region    /note="MER30 element fragment"
                 20661..22730
repeat_region    /note="MLT1B element fragment"
                 22740..22814
repeat_region    /note="MLT1A element fragment"
                 22917..22974
repeat_region    /note="MLT1A element fragment"
                 23342..23501
repeat_region    /note="THE1B element fragment"
                 23550..23697
repeat_region    /note="MSTC element fragment"
                 23725..24279
repeat_region    /note="THR element fragment"
                 24395..24593
repeat_region    /note="THR element fragment"
                 24643..25270
repeat_region    /note="THR element fragment"
                 25277..25463
repeat_region    /note="THE1B element fragment"
                 25528..25627
repeat_region    /note="THE1B element fragment"
                 26001..26196
repeat_region    /note="MLT1A element fragment"
                 26265..26661
repeat_region    /note="LTR2 element fragment"
                 26697..26898
repeat_region    /note="MSTC element fragment"
                 27949..28241
                 /partial
repeat_region    /note="Alu repeat: matches 308.1 of consensus"
                 28319..28608
                 /partial
repeat_region    /note="Alu repeat: matches 308.1 of consensus"
                 30032..30159
                 /partial
repeat_region    /note="MER5 element fragment"
                 31794..31980
                 /partial
repeat_region    /note="Alu repeat: matches 308.109 of consensus"
                 31981..32073
                 /partial
repeat_region    /note="Alu repeat: matches 93.1 of consensus"
                 33232..32735
repeat_region    /note="MER9 element fragment"
                 35168..35654
repeat_region    /note="MER4B element fragment"
                 40640..40693
repeat_region    /note="27 copies of 2 mer 94 % conserved"
BASE COUNT       11544 a    7879 c    8141 g    13150 t
ORIGIN
```

```
Qy    688 ytncaycaytcgyaargtnaayathtggytnatgmgnaaarcarathytngcnaayaaygar  747
          |||:||||||| ||||||  ||||||::::||| ||||:| ||::||||| |:::||| :||
Db  24003 ctgccctagagatctgtggaattgacctgagacgagatgaattaggtatcaggtggaa  24062

Qy    748 garathwsnaarcarwsnathcaargaargtnacntgggtnytnytnaargcnttywsn  807
          :||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24063 gaaatttctaagcacacaaaagcattcaagagagtgactgttgtgtccttgttaaaggcacactcagt  24122
```

```
Qy    808 ttyathngargtcngnarcayaarwsbwsngaraaytncayccngayaaygtnathaar  867
          ||:|||||||||||||||||||||||||||||||||||||||||||||||||||:||:|
Db  24123 TTTGCAAGGGAGCCAGAACATAAAAGTTCAGAAAATTTTCAGCCTGACAAATGTGATAGAA  24182

Qy    868 aaraaaraaycc-nttywsngargnaartty aarytngcngnagarathtgyathtgyaa  926
          || ||| |||| ||:||||||||| |:|||| ||||||||| |||||||:|||||||||:|
Db  24183 AAGAAAAACCCATTTTTCTGAGATGAAATTAGGTATCGAGAAAATTTGCATAAGTAA  24242

Qy    927 ygargarytnaaygtnaayccncaargayayayggngaraayaythwsntggactngyacm  986
          |:||:|||| ||||||||| |::|||||||| |||||||||||||||||||| |||| |||
Db  24243 CCGAGGAGCCAAATGTTAATCACTAAGACCATGGGGAAAAATATCTCCATGGCTTGTCAGAG  24302

Qy   1046 gnacngnccnggnwsnytntcgytgyttcancccnmgagayyytngtncCntgygytnccng  1105
          ||||||||||||||||| |:||| ||||||||||||||||||||||||||||||||||||||
Db  24303 GTCTTCACGGCAGCCCTCACATCACTCCCCCCAGAGCCTAGGAGAGAAAATTGTTTTATG  24362

Qy   1106 tnaa------------------------------ywsngcngtngcnwsngaarggng  1132
          |||:                             :|| |||||||||||||||||||||
Db  24363 GGTCGGGCCGCCCAGGGTCCCCATGCTATCTGTGCAGCTCTAGGGACTTGGGTGTCCCAG  24422

Qy   1133 cnwsnccnaaarccntggcaarytnccnwsnggngtngarccngtngtggngcnaaraarwsnm  1192
          |||| || |||||||||||:||||||| |||||||||||||||||||||| |||||||||||||
Db  24423 CCACTCCAGCCATNAACTGAAAAGGGGCCCAAAGGTATAGCTTGGGGTTTGCTTCAAAGGATG  24482

Qy   1193 gnathgargtntggarccnccnccnathgnttycararaarathtaygganaaycchtgatgc  1252
          |:|||||| ||||||||| || ||||||||||||||||||| |||||| |||||||||||:|
Db  24483 GAAGCCCCAGGCCTTGGCGAGCTTCCATGTGGCGTTGAGCCTGCAGGTGGCCACTGAAGTCAA  24542

Qy   1253 cnmgncaraarttygcngtngnggngtngghwsnsntggmgnacnwsngcnmgngtngtnc  1312
          |:|| |||||||||||||||:|||||||||:||||||||:||||||||||| ||| |||||
Db  24543 GAATTGAGGTTGGGAACCTCACCCTAGATTTCAGAAGAGTGTATGGGAATGGCCTGGGATGC  24602

Qy   1313 araaargnnaaaytnggntnggnarccncncncacaymgngtnccnwsnggngcnconwswsnsnm  1372
          ||||||| || ||||||||:|||| |||:| ||||||||||| ||||||| ||||||:|| :||
Db  24603 CCAGGCAAAAAGTTTGCTCCAGGGGGGGGGAGGGACCTCATGGGAGAACCTCTCGTGGGTGT  24662

Qy   1373 gngcngtnamgngmgnwsncccncnwsnsmgnytncarraarggnamgnwsnacngaywsny  1432
          || ||||||:|||||||| |||||||:|| :|||||||||||||:|||||||| |||:||||
Db  24663 GGAAGGGGAAAAGTGGGGTCATAGCCTACCCACACAGAGTCCCTACCAGGGGACCACCTAGTG  24722

Qy   1433 tncarcayytncngaraarwsnacngayacncarytgycarccngtnaargcngncnggna  1492
          ||||||| ||||||||| || |||||| ||||||||||||||||||||||| |||||||||
Db  24723 GAGCTGTGAGAGAGGGGCCCACCATCCTCCAGACCCAGAATGGTAGGATCCCAAGATGGTCA  24782

Qy   1493 tggarwsngtncntayaaracngtngtngcngaryttnacnaaracngtngnathtaty  1552
          ||||| ||||:|||||| ||||||||||||| ||||:|| ||||||||||||||:||||
Db  24783 TGCACCCGTGTGCCTGGAAAAGCCCGACAGCAACTCAAATGCCAGTCTGTGAAAGCAGGA  24842

Qy   1553 tnyttncaytgycaygayytngayvtgncayggngtnaaргtngaycayttyggngncy  1612
          |||||||||||:||||||||:|||:|| ||||||||||||||||||| ||||||||||:
Db  24843 GGGAGGCGTGTACCCTGCAAAGCCACAGGGCCAGAGATCATGGGAACCAATC  24902

Qy   1613 tnmgnttygaytgyccnacncnggnttymgnacntayatgggncongt-ncnyntgtyty  1671
          |||||||||||||||:|| |||||||||||:|||:|||:||| |||| :||||:|||:||
Db  24903 TCTTGCATCAGCATGACCTGGATATGAGACAGTCAAAAGGCAAACATTTTTGGACTT  24962

Qy   1672 ggncaartytyccntyggnacngcngtnttcacncaartgyyttntavytncaytgyat  1730
          ||||||||:||||||:|||||||| |||:|||||||||||||::|||||:|||:||||:|
Db  24963 TAAGATTTGACTGTCCCACTGGAATCTCAGACTTCCATGGGGCCTGTGTACCCCATTGTAT  25022

Qy         
Db  25023 GGCCAAATATCTCCACTGGAATGGGCTGTATTTATCCAAATGCCTGTACCCCATTGTAT  25081
```

```
Qy    688 ytncaycaytcgyaargtnaayathtggytnatgmgnaaarcarathytngcnaayaaygar  747
Db  24003 ctgccctagagatctgtggaattgacctgagacgagatgaattaggtatcaggtggaa  24062

Qy    748 garathwsnaarcarwsnathcaargaargtnacntgggtnytnytnaargcnttywsn  807
Db  24063 gaaatttctaagcacacaaaagcattcaagagagtgactgttgtgtcctTGTTAAAGGCACACTCAGT  24122
```

us-09-997-610-7.rge

```
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB 14 - 14 DNA,
            clone_lib:Keio BAC library clone:KB1222D11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159712)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     On Jan 14, 2002 this sequence version replaced gi:15824050.
FEATURES             Location/Qualifiers
     source          1..159712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1222D11"
                     /cell_line="FLEB 14 - 14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     repeat_region   1..4242
                     /rpt_family="L1PA2"
                     /evidence=not_experimental
     repeat_region   4668..4733
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   5899..5921
                     /rpt_family="(TGGA)n"
                     /evidence=not_experimental
     repeat_region   complement(6205..6406)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   6410..6589
                     /rpt_family="FAM"
                     /evidence=not_experimental
     repeat_region   6619..6648
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(6649..6776)
                     /rpt_family="AluJo/FRAM"
                     /evidence=not_experimental
     repeat_region   6860..7158
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   9083..9127
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   9188..9226
                     /rpt_family="FRAM"
                     /evidence=not_experimental
     repeat_region   complement(9227..9522)
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   9836..9933
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   10728..10762
                     /rpt_family="L1MA1"
                     /evidence=not_experimental
     repeat_region   10763..11062
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   11063..11300
                     /rpt_family="L1MA1"
                     /evidence=not_experimental
     repeat_region   11301..11324
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   11520..11545
                     /rpt_family="AT_rich"
```

```
     repeat_region   12507..12560
                     /evidence=not_experimental
     repeat_region   12793..12825
                     /rpt_family="(TAAAAA)n"
                     /evidence=not_experimental
     repeat_region   15792..15852
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   15993..16013
                     /rpt_family="AluJ/FRAM"
                     /evidence=not_experimental
     repeat_region   complement(16014..16312)
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   17126..17436
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   18198..18489
                     /rpt_family="AluSc"
                     /evidence=not_experimental
     repeat_region   complement(17798..17960)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   18007..18197
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   18490..18715
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(18801..18936)
                     /rpt_family="FLAM_C"
                     /evidence=not_experimental
     repeat_region   complement(19119..19294)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   19346..19541
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   19687..19844
                     /rpt_family="L1MEc"
                     /evidence=not_experimental
     repeat_region   complement(19954..20081)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   20082..20106
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   20423..20566
                     /rpt_family="FRAM"
                     /evidence=not_experimental
     repeat_region   20567..20636
                     /rpt_family="GA-rich"
                     /evidence=not_experimental
     repeat_region   complement(20973..21282)
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   22379..22724
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   22933..23019
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   23044..23270
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(23275..23586)
                     /rpt_family="AluSg1"
                     /evidence=not_experimental
     repeat_region   complement(23946..24035)
                     /rpt_family="MER94"
                     /evidence=not_experimental
```

```
repeat_region    24491. .24777
                 /rpt_family="AluY"
                 /evidence=not_experimental
repeat_region    24782. .24794
                 /rpt_family="AT_rich"
                 /evidence=not_experimental
repeat_region    24795. .24909
                 /rpt_family="AluJb"
                 /evidence=not_experimental
repeat_region    24910. .24939
                 /rpt_family="LIPA4"
                 /evidence=not_experimental
repeat_region    24910. .24939
                 /rpt_family="AT_rich"
                 /evidence=not_experimental
repeat_region    complement(24981. .25276)
                 /rpt_family="AluJb"
                 /evidence=not_experimental
repeat_region    complement(27879. .28173)
                 /rpt_family="AluSp"
                 /evidence=not_experimental
repeat_region    complement(28598. .29054)
                 /rpt_family="MER63B"
                 /evidence=not_experimental
repeat_region    30710. .30738
                 /rpt_family="AT_rich"
                 /evidence=not_experimental
repeat_region    complement(30739. .30993)
                 /rpt_family="AluJb"
                 /evidence=not_experimental
repeat_region    30994. .31032
                 /rpt_family="AT_rich"
                 /evidence=not_experimental
repeat_region    complement(31408. .31724)
                 /rpt_family="AluJo"
                 /evidence=not_experimental
repeat_region    complement(31949. .32254)
                 /rpt_family="AluSg"
                 /evidence=not_experimental
repeat_region    32336. .32362
                 /rpt_family="AT_rich"
                 /evidence=not_experimental
repeat_region    complement(32463. .32565)
                 /rpt_family="L2"
                 /evidence=not_experimental
repeat_region    32685. .32898
                 /rpt_family="MIR"
                 /evidence=not_experimental
repeat_region    32995. .33020
                 /rpt_family="(TAAA)n"
                 /evidence=not_experimental
```

```
Query Match
Best Local Similarity   48.0%;
Matches  484;  Conservative 210;

Qy    744 rgararathwsnaarcaarwsnathcargargtnacntgggtnytnaytnaargcntt 803
          :|:||::||:.:||::| ||:||:| |||:|.|:|::||::| ::||||| |:.|| |::||
Db 138232 AGAGAAGTTCCTAAGCAGCAAAAGCATTCAAGAGATAACTTGGGTGTCATTAAAGGCACT 138173

Qy    804 ywsnttyathmgnsgargcngarcayataaarwsnwsngaraayytncaycngayaay 863
          ::: |||||:| ||::|| |:|| ::|::||:|:| |:|:: |||:|.::||::|:|:|
Db 138172 CAGTTTTATAAGGGAGCAGAGCATAAAACTTTGGAAAAACTTTCGCACCTGACTATCCAAT 138113

Qy    864 haaraaaraaaycnttywsngargngaaarttyaaarytngathtgyathtg 923
          :: |||:| ||::||| ||:|:| ||:|| :||||:| |:.|::||:| |||:
Db 138112 AGAAAAAGAAAAAACCCATTTTCTGGGGAGAAATTCAAGCCGGTCAGAAAATTTGCACAAG 138053

Qy    924 yaaygargaryytnaaygtnaayccncaargayaaayggngaraaathwsntggactnyca 983
          :|:||::|:: ||::||:|::|||:|||| | ::|| :|::|||:|| ::||||:||
Db 138052 TAGCAAGGAGACCAAATGTTAATCCCAGGACCATGGGAAAATGTCTCCAGGCCATGTCA 137993

Qy    984 rmgnvwsncarcaarwsnathaarwsnytngcntgggmgnc---cnmgmgnaartggtt 1040
          :-| :|||::  |||:| ||:.:: ::.||: ||:||||: ||.||::|||:||||||
Db 137992 GAGACCTTTACAAGCAGCCCCTCCATCATAGCCTGGAGGCCTACCAGGAAAAGTGGTT 137933

Qy   1041 ytgyggnacngngccnggngnwsnytntgtygytncarccnmngayytngtnccntgygt 1100
          ::|:||| :| :||:.||:|:||::||:||| ||:|||:::|::|:| ||:|:.|| ||:
Db 137932 TCATGGGCCTGGCCTGGCCGCTGGTCACGCCTAGGGACTTGGGACCTGTGACCTGTGT 137873

Qy   1101 nccngtnaaywsngcngtngcnwsngargg----------ngcnwsncnaa 1142
          ::| |||:.|:||:||:||:|| ||:|:          ||:||:||||
Db 137872 CCCAGCCACTCCAGCCGTCGGGCTGCTGAAAGGCGTACAGGGGCTGGGTGTGATTGCTTCAGA 137813

Qy   1143 rcontggcaryytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnmathgargt 1202
          | ::|:|| |::|| |||:||:||| ||:||:||:|:||::|:|::| :|:|| .||:|:
Db 137812 GGCTTGGCCAGCTTCACCTGGTGCTTGACCCTGGGTGCCACAGAAGTCAAGAATTGAGGT 137753

Qy   1203 ntgggacccncccnaathmgnttycaaraaarathtayggnaayccntggatgccnmgncaraa 1262
          ::|||||::|:::| |:|:||:||:::|||:|:||||:|:|| |:|::||::|| |:|::||
Db 137752 TTGGGAACCTCTGCCTAGATTCAGAAGATGTTATGGAAAATGCCCAGGCAAAA 137693

Qy   1263 rttygcngtngggtnggnwsnwsntggmgnacnwsngcnmgntngtncaaraargnaa 1322
          :||:||:||||:||:||::|:|:|:||::|::|::||:|| |::| ::|| |||::
Db 137692 GTTTGCTGCAGGTGGGGCCCTCATGGAGACCTCTGCTAGGGCACTGAGGAAGGAAAA 137633

Qy   1323 ygtnggngtgggarccncccncaymgngtncccnwsnggngtncccnwsnwsnmgncngtnmg 1382
          :|||||||:||::||::|::|:||::|.||:||::|||||::|.:|::|:|:|:|:| ||
Db 137632 TGTGGGGTTCGATTCCCCACACAGAGTCCCTACTGGGGCCACTGCCTAGTGGAGCTGTGAG 137573

Qy   1383 nmgnwsncnccnccnwsnwsnmgnytncaaraarggnmgnwsnaacngaaywsnytncaaraygt 1442
          ::|||:|:||||:||:| ||:||::||:||:|:||:::|:|:|||:||:::||:|||::
Db 137572 AAGAGGGGCCACCATCCTTCAGACCCCAGAATGGTAGATCCACTCGACACATGT 137513

Qy   1443 nccngaraarwsnacngayacncaartgycaarccngtnaargncngcnggnatggaarwsngt 1502
          :||| ||:|||::|:||||:|:| :||||:|:||||::||:|::|:||:|||:| |:||::|:
Db 137512 GCCTGGGAAAAACTGCAGACATTCAATGCCAGCCCATGAAAGCCAGGAGGCTAT 137453

Qy   1503 nccntayaarcangtngtngcngaryytnacnaaracngtnggnaathtayytnytncaytg 1562
          ::|:||:||:| ||:| |:|| |::||:|||:||:|:|| :|::|||:|.|::||:||:||
Db 137452 ACCCTGCAAAGCCACAGGGCAGAGCTCCCAAGACCATGAGAACCTACCTTTGCATCA 137393

Qy   1563 ycaygayytngayytngaygyggngtnaaramgngaycayttcyggngncnytnmgnttyga 1622
          :|| ||:|||||:|||||:| |::|||:| ::|::|||::||:|:|||:|.:|:|:||:|::
Db 137392 GCCTGACCTGGATGTGAGACCTGTAATCAAAGAGATCATTTTGGAGCTTTAAAATTTGG 137333

Qy   1623 ytgyccnacngngnttymgnacntayatgggnccngt-nccnytntgyttcyggncarttyt 1681
          :||:|:||:||:|| |:|||:||:||||::||::||| :|||:|||:|:|::||||:|::||
Db 137332 CTGCCCCACTCGTTTTTGGACTTCCATGGGCCCGTGTAACCTCTTTGTTTTTGGCTAAATTTT 137273

Qy   1682 tyccnttcyggnacngcngtnttyacncaartgyytntayytncaytgyat 1730
          :|||:||| |||:| ||:| |||:.||:|||:|:|::||:|||::|:|||
Db 137272 TCCCATTTGGAACAGTACGTTTCCCAATACCTGTACTCCCATTGTAT 137224
```

```
Score 487.6;  DB 9;     Length 159712;
Pred. No. 5.5e-108;
Mismatches 293;  Indels 22;  Gaps 3;
```

```
Search completed: July  8, 2002, 18:27:45
Job time: 19721 sec
```

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: July 8, 2002, 16:20:51 ; Search time 481.64 Seconds
                                              (without alignments)
                                              6170.540 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg.........gyytntaytcaytcgyatg 1731

Scoring table: IDENTITY_NUC
                       Gapop 10.0 , Gapext 1.0
Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_032802:*
        1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
        2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
        3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
        4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
        5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
        6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
        7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
        8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
        9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
       10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
       11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
       12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
       13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
       14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
       15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
       16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
       17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
       18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
       19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
       20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
       21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
       22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
       23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
       24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB ID	Description
1	471.4	27.2	2590	22	AAH14327	Human cDNA sequenc
2	452	26.1	3093	22	AAS03687	Rhesus gene locus:
3	452	26.1	9236	22	AAS03689	Rhesus gene locus:
4	452	26.1	9236	22	AAS03690	Rhesus gene locus:
5	452	26.1	9241	22	AAS03688	Rhesus gene locus:
6	429.8	24.8	3203	22	AAD05134	Human secreted pro
7	427.6	24.7	6063	22	AAL03635	Human reproductive
8	427.2	24.7	22509	22	AAS44505	Human LEKTI DNA c1
9	427.2	24.7	31529	22	AAS44506	Human LEKTI DNA c1

10	426.4	24.6	6063	22	AAL03634	Human reproductive
11	421.4	24.3	1736	22	AAH13678	Human cDNA sequenc
12	421	24.3	465237	24	ABA90193	Human oestrogen re
13	409	23.6	1485	23	AAS91984	DNA encoding novel
14	408.8	23.6	1278	22	AAD05178	Human secreted pro
15	400.4	23.1	1580	23	AAS87262	DNA encoding novel
16	395.2	22.8	7726	22	ABA21079	Human nervous syst
17	393.6	22.7	32127	22	AAI99255	Human excretory re
18	393.6	22.7	32127	22	AAI63605	Human kidney relat
19	393.6	22.7	72215	22	AAK868 32	Human immune/haema
20	388.2	22.4	1405	22	AAK70266	Human immune/haema
21	386.6	22.3	1405	22	AAK70265	Human immune/haema
22	386.4	22.3	8387	22	AAK84892	Human nervous syst
23	386.4	22.3	8387	22	AAK84548	Human immune/haema
24	384.4	22.2	1621	23	AAS67924	DNA encoding novel
25	378.8	21.9	3049	22	AAA97997	Human T gene DNA.
26	373.8	21.6	81369	21	AAH16637	DNA encoding human
27	371.6	21.5	1306	23	AAS403 21	DNA encoding novel
28	361.6	20.9	22756	22	AAL03921	Human reproductive
29	361.6	20.9	22756	22	ABA44227	Human breast cell
30	355.4	20.5	1946	22	ABA54677	Human foetal liver
31	355.4	20.5	1946	22	ABA24459	Human brain expres
32	355.4	20.5	1946	22	AAK02964	Human bone marrow
33	355.4	20.5	1946	22	AAK28410	Probe #2908 for ge
34	355.4	20.5	1946	22	AAI12975	Probe #2920 used t
35	355.4	20.5	1946	22	AAI34334	Probe #3020 used t
36	355.4	20.5	1946	22	AAI02893	Probe #2884 used t
37	355.4	20.5	26410	22	AAK70623	Human immune/haema
38	347	20.0	1946	22	AAK69758	Human immune/haema
39	346.8	20.0	32986	22	AAK84629	Human immune/haema
40	346.4	20.0	32986	22	AAS66623	DNA encoding novel
41	346.4	20.0	923	23	AAD26400	Human glutamate re
42	341.4	19.7	50000	24	AAD26437	Human GRM3 gene fr
43	341.4	19.7	50000	24	AAC87588	Human 9p11 chromos
44	337.4	19.5	57728	22	AAH18389	Human cDNA sequenc
45	336.6	19.4	2197	22		

ALIGNMENTS

RESULT 1
ID   AAH14327 standard; cDNA; 2590 BP.
XX
AC   AAH14327;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11697.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.

```
XX
PT  Primer sets for synthesizing polynucleotides, particularly the 5602
PT  full-length cDNAs defined in the specification, and for the detection
PT  and/or diagnosis of the abnormality of the proteins encoded by the
PT  full-length cDNAs -
XX
PS
XX
PT  Claim 8; SEQ ID 11697; 2537pp + CD ROM; English.
XX
CC  The present invention describes primer sets for synthesising 5602
CC  full-length cDNAs defined in the specification. Where a primer set
CC  comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC  to the complementary strand of a polynucleotide which comprises one of
CC  the 5602 nucleotide sequences defined in the specification, where the
CC  oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC  of an oligonucleotide comprising a sequence complementary to the
CC  complementary strand of a polynucleotide which comprises a 5'-end
CC  sequence and an oligonucleotide comprising a sequence complementary to a
CC  polynucleotide which comprises a 3'-end sequence, where the
CC  oligonucleotide comprises at least 15 nucleotides and the combination of
CC  the 5'-end sequence/3'-end sequence is selected from those defined in
CC  the specification. The primers can be used in antisense therapy and
CC  in gene therapy. The primers are also useful for synthesising polynucleotides,
CC  particularly full-length cDNAs. The primers are useful for the
CC  detection and/or diagnosis of the abnormality of the proteins encoded by
CC  the full-length cDNAs. The primers allow obtaining of the full-length
CC  cDNAs easily without any specialised methods. The primer sets. AAH03166 to AAH13628 and
CC  AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC  AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC  represent oligonucleotides, all of which are used in the exemplification
CC  of the present invention.
XX
SQ  Sequence 2590 BP; 722 A; 526 C; 670 G; 672 T; 0 other;
```

```
Query Match        27.2%;   Score 471.4;   DB 22;   Length 2590;
Best Local Similarity 47.0%;   Pred. No. 2e-110;
Matches 481;  Conservative 213;  Mismatches 293;  Indels 37;  Gaps 3;

Qy  744 rgargarathwsnaarcaarwsnathcaargargtnactnggtnytntnaargcntt 803
         |||:|||:||::|| ::  ||||:||||||||||:||:|: :|| ||||| |
Db  608 agaagaaatttctaagcagcaaagacattcaagaggtgagtggtactgttaagagcatt 667

Qy  804 ywsnttyathmgngargcngarcayacavyarwsnwsngaraayayyccngayaavygtnat 863
         ||:||| |||||:|||||:||:|: |  :|| ::| ||||::|||:|| ||||||:|
Db  668 cagttcaaaggaaacagcataaaaattcagaaaatttgtagcctgacttgcaat 727

Qy  864 haararaaraayccnttywsngarggnaarttyaaryntngcngnaarathtgyathtg 923
         ||:| ||:|||:::|  :|| :||||:||||| ||||:  ||||:||:|||::||:
Db  728 agaaaaagaaaaccagtttctgggagaaatttcaagccagctgcagaaatctgcataag 787

Qy  924 yaaygargartytnaaygccncaraayaayggngaraaayathwsntggactngyca 983
         :|| |||||: :||::||| ::| ||||:|||:||||| |||:||:||||| |:||
Db  788 catcaagaagcataatgttaatcccaggaaccatggggaaaatgtctccaggccaatgtca 847

Qy  984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnng---naartggvtt 1040
         :| ::| ::|||||||::|||  :||:|||||||::|:||| ||:||   ||||||||
Db  848 gagacgcttcacagcagccccctccctgtcacagcccagagatccagagggaaaaaatggtt 907

Qy 1041 ytgyggnacngngnccnggnwsnytntgytgyngtncaarccnmgnayaayytngtncn 1100
         :||:||:|| |:|::| ||::|:||:|||||:||:|||||||| ||:|:|| ||||:|
Db  908 ttatggcaggaccaagggtccctgtgctgtgtgcagcctatagaacttggtgccctgtgt 967

Qy 1101 nccngtnaayws-------------------ngcngtngcwsnga 1127
         :|||||||:| |                  :| ||:|| ::|||
Db  968 ccagctgctccagcaccatggctgaaaggggccaatgtcatgctggcttcaga 1027

Qy 1128 rggngcncnwsncnaarccntggcarytncnnwsngngngtngarccnmgnggncnaaraa 1187
         :||:||:|:::|||||||::|||||::||::| :|:|:|:||||||:||:||:| ||||||
Db 1028 gggtggaagcccaagccttcacatggtgttgagcctgtaggtacacagaa 1087

Qy 1188 rwsnmgnnathgargtntggarccncnncnnathmgnttycaraaraththayggnaayccntg 1247
         :::||:||:||||:|:||||:||:|:|:|:|||:||:|||||||||:|:||||::|||:||:|:
Db 1088 gtcaagaaattgaggtttgggaacttccacctagatttcagaagagaatgtatatggaaatgcctg 1147

Qy 1248 gatgccnmgncaraarttygcncngtngggntnggngnacnwsngcnmgngt 1307
         ||||:|:||:|||||:||:||:|:||:|||:|:||:|:|||::|||:||:||
Db 1148 gatgtccaagcaaaagtttgctgcaggtggcaggcgcctcatggagagaatctggctaggggc 1207

Qy 1308 ngtncaaraarggnaaytgngntggggccngagctcccacgtagaagtcccacqtnaargcngc 1367
         :||:||||||:||:||:||:|:|||||||:||||||||:||:|||:|||:|::|:||||||:|:
Db 1208 aatgcagaagggaattgtggggcggagctcccacgtagaagtcccacgtaaatgggcattatc 1267

Qy 1368 nwsnmgngcngtnmgngmgnwsnccncnwsnwsngnytncararggngmgnwsnacnga 1427
         :::||:||:||:|:|:|::|:::|||:|:::::|:|:||:|||:|:|:|::|:::|:||
Db 1268 tagtggagctgtgagaagaggcccaccgtccttcagacccacagtggatccactga 1327

Qy 1428 ywsnytncarcaygtncngararwsnacngayacnayctgycarccngtnaargcngc 1487
         :::|:||:|||:||:|:|:||::::|||:||:|||:|:|:||:|||:|:||||:||:|
Db 1328 cagcttgcaccatggcgcctggaaaagccgcagacaactcatcggctgaaaagca 1387

Qy 1488 nggnatggarwsngtnccntayaaracngtngtncgngarytnacnaaracngtnggnat 1547
         :|||||||:::|:||:||:||:||:||:||:||:|:|||:||:||:|||||:||:||||
Db 1388 caggagggagggctgtaccctgcaaggcacagggtccgcaaaagccatggggtcc 1447

Qy 1548 htayytnytncaytgycaygayytngaygngtnaarmgngaycayttygg 1607
         :||:||:||:||:||:||:||:||:||:|:||||:|:|:||:||:||:||
Db 1448 ccacctttgtcatcagcgtgacctgaatgtggacctagagtcaaaggagatcaatttg 1507

Qy 1608 ngcnytnmgnttygaytgyccnacnggngnttymgnacntayatggngnccng-tncnytnt 1666
         :||:||:|:||:||:||:||:|||:|:|:||:|:||||:||:|||:|:||||| :|:|:||:
Db 1508 agcttaaaattgattgtcctgctgatttggagctggtcctgcagccctt 1567

Qy 1667 gyttyggncaarttyttyccnttyggnacngcngtnttyacncaartgyytntaayttncayt 1726
         |:||:||:||||||:||:||:||:||:|||||:||:||:||:|||:||::|:|||:||:||:
Db 1568 gtttgtgccaatgtctccatttggaatggctgtatttacccaataacctgtaccctatc 1627

Qy 1727 gyat 1730
         |:||
Db 1628 gtat 1631
```

```
RESULT 2
AAS03687/c
ID   AAS03687 standard; DNA; 3093 BP.
XX
AC   AAS03687;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Rhesus gene locus: RHD gene deletion in Rh negative haplotypes.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   primer_bind    32..54
FT                  /*tag= a
FT                  /note= "Binding site of primer rez7"
FT   primer_bind    complement (3034..3054)
FT                  /*tag= b
FT                  /note= "Binding site of primer rnb31"
XX
PN   WO200137702-A2.
XX
PD   31-OCT-2000; 2000WO-EP10745.
XX
PF   10-MAY-2001.
XX
PR   02-NOV-1999; 99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
```

PI Flegel WA, Wagner FF;
XX
DR WPI; 2001-291052/30.
XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX
PS Example 10; Fig 5; 135pp; English.

CC The sequence represents the coding sequence of Rhesus gene locus:
CC RHD gene deletion in Rh negative haplotypes. The Rhesus genes
CC locus comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;

Example 10: Fig 5; 135pp; English.

Query Match 26.1%; Score 452; DB 22; Length 3093;
Best Local Similarity 47.4%; Pred. No. 2.4e-105;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

Qy 744 rgargarathwsnaarcarcarwsnathaargargtnacntgggtnytnytnaargcntt 803
         : |||:|:::| | | |||| |:|||:|: || | | ||||::|| :
Db 1338 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 1279

Qy 804 ywsnttyathmgnaargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
         : ||||:|:| |||:|| |:|||:||: || :| || | |:||| ||| |
Db 1278 CAGTTTTATAAGGAAGCAGAGCATAAAAAGTTCGGAAAATTCCAGCCTGACAATGTGAT 1219

Qy 864 haaraaaaaraa--ycnttywsngarggnaattyaarytngcngcngarathtgyath 921
         : ||||| |||:|| |:|||||||| | : |||| ||||::| |||:|| |
Db 1218 AGAAAAGAAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGCTGCAGAAATTTGCATG 1159

Qy 922 tgyyaaygargaryytnaaygtnaayccncaargayaayggngaraayathwsntggacntgy 981
         | ||| | || || |:|| :| | :|| || ||:| ||| | | :||:
Db 1158 AGTAAC-AGGAGCAAAGCTAATTCCCAAGACAAATGGGAAAAATGTCTCCAGGGCATGT 1100

Qy 1158 AGTAAC-AGGAGCAAAGCTAATTCCCAAGACAAATGGGAAAAATGTCTCCAGGGCATGT 1100

Qy 982 caringnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgmgnaartggtt 1040
         : | || ||::|: |:||| ||| ::|| || ||| ||||| ||||| | |||||:||
Db 1099 CAGAGGTCTTTATGGCAACCCCTCCATCACAGTTCCAGAGTATCAGGAAAAATGGTT 1040

Qy 1041 ytgyggnacngggnccnggngawsnytntgytgyrgtncaarccnmgnnayytngtncntgyyt 1100
         :|:| ||||| |||| |::|:|:|||::|| ||||:|||:||||| |:|||
Db 1039 TTGTTGGCCAGGCCCGGCGTCCCTCATGCTGTGTCCAGCCTAGGGACTTGGTGCCCTGCAT 980

Qy 1101 nccngtna------------------aywsngcngtngcnwsng 1126
         |||:| | :: ||||| || | |:|::| |
Db 979 CCCAGCCACTCCCAACCATGACTGACGGGGGCAAAGGTAGAGCTTGGGCTGTAGCTTCGG 920

Qy 1127 arggngcnwsnccnaarccntggcarytncccnwsnaggntngarccngtnaggngcnaara 1186
         |:|| | |||::|| |:|:| ||:| |:| |||: | :||| || |
Db 919 GGAGTCGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTTGAGACTGCAGAGTCCACAGA 860

Qy 1187 arwsnmgnathgargtntgggarccnccnnaathmgnttycararaththtayggnaaycent 1246
         |:::| | |||| | || | ||| | ||| |||: :| || | |:: |||
Db 859 AGTCAAGAAGATTGGGGTTTGGAAACCTTCCCTAGATTAAAGAGGATGTGTCCGGAAATGCCT 800

Qy 1247 ggaatgccnmgncaaraarttycgncgtnggngtnggnwsnwsnttcggmgnacnwsngcnmgng 1306
         |||||:|| ||| | |:|:||||::|||||||:|:||||:|| || ||::|| ||||:| |
Db 799 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGAGCAGGGCCCTCATCAGGATCCTCTCCAGGG 740

RESULT 3
ID AAS03689/c
AC AAS03689 standard; DNA; 9236 BP.
XX
AC AAS03689;
XX
DT 29-AUG-2001 (first entry)
XX
DE Rhesus gene locus: upstream Rhesus box of D-positives.
XX
KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD;
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS Homo sapiens.
XX
PN WO200132702-A2.
XX
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000WO-EP10745.
XX
PR 02-NOV-1999; 99EP-01216686.
PR 31-MAY-2000; 2000EP-0111696.
XX
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI Flegel WA, Wagner FF;
XX
DR WPI; 2001-291052/30.
XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX
PS Disclosure; Fig 9; 135pp; English.

Qy 1307 tngtncaaraorggnaaygtnggntgggaarccncnccnncaymgngtncccnwsnaggngngcnccnw 1366
          |||:|| ||||||::|||||:: ||||| |:||||| |||||||
Db 739 CAGTGCAGAAGGGAAAAATGTGGGGTGGGGTCAGAGACACACACAGTCCTTACTCGGGGCACCAC 680

Qy 1367 snwsnmgncngtnmgmgnwsnccncnwsnwsnmgnytncnaraargngmgnwsnacng 1426
          || :| || | || |::| ||:||||:|| || |:||||::|:| ||| ||
Db 626 ACAGCTTGCACCGTGTACCTGGGACAAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 567

Qy 1427 aywsnytncarcaygtnccngararwsnacngaayacncartgtycarccngtnaargcng 1486
          ||:| | || |:||| :||| |:|| | |:| |||: ||: |||||||::|:|:|| |
Db 566 CTGAGAGGGAGGGCTGTACCCTGCGAAAAGTTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 507

Qy 1487 cnggnatggarwsngtncctayaaracngtngtngcngarytnacnaaracngtnggna 1546
          ||| ||||:| |||:: ||||||||:||||:|| :||||:| |||||||||:||| |
Db 506 CCCAACCCCTTCCATCTGCGTGAACCTGGATGCTGGAGTCGAAGGAGGATCATTTTG 447

Qy 1547 thtayytnytncaytgtycaygayaytngaygtnmgncaygggntnaarmgngaycayttg 1606
          |:| |:|:| | | |::| |:|| |:|| || : | |::||||::|||||| | |
Db 506 CCCAACCCCTTCCATCTGCGTGAACCTGGATGCTGGAGTCGAAGGAGGATCATTTTG 447

Qy 1607 gngcnytnmgnttygaytgyccnacngngnttymgnacntayatgggnccngtnccnytnt 1666
          ||||:|| | || ||| | | | |: :|||||| ||:::|||
Db 446 GAGCTTTAAGATTTGACTGCGCCCCACTGACATTTTGGCATCCATGGGCCTGTAGCCTCTTT 387

Qy 1667 gyttyggncarttytyccnttyggnacngcngtnttyacncaartgyytntaytytncayt 1726
          |:| |||| |::||:| | || || |:|||||| | |:| ||| | |:::|:|
Db 386 GTTTGACCAATTTATCCCATTTGGAATGCCTGTATTACCCAATGCCTGTACCCCATT 327

Qy 1727 gyat 1730
          :|||
Db 326 GTAT 323

XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX
PS Disclosure; Fig 9; 135pp; English.

CC The sequence represents the coding sequence of Rhesus gene locus:
CC upstream Rhesus box of D positives. The Rhesus genes locus

CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;

Query Match   26.1%; Score 452; DB 22; Length 9236;
Best Local Similarity   47.4%; Pred. No. 6.3e-105;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

Qy   744 rgargarathwshnaarcaarwsnathcaargargtnacntgggtnytnaargcntt 803
          :|:||||:|||:||||||| ::|:||| | |||| ||:||||||:||| || |||
Db   6014 AGGAGAAATTTCTPAAGCACGCAAAAGCCATTCAAGAGGTGACTGGGTCGTTAAAGACATT 5955

Qy   804 ywsnttyathmngngargcngarcayaarwsnswsngaraayytncaycycngaayygtnat 863
          :|| |||:|||||| ||| |||||| |||:|||||||||||| | || ||||| ||| ||
Db   5954 CAGTTTTATPAAGGGAAGCAGACAATAAAAAGTTCGGAAAAYTTGCACGCCTGACACAT 5895

Qy   864 haaraaraaraaa--ycenttywsngargnaarttyaarytngcngncngaarathtgyath 921
          |:|||:|||| || ::|||| ||||:|||||||||| ||| ||| || ||||| ||||
Db   5894 AGAAAAAGAAAAATTCCCATTTTCTGAGGAGAAAATTCAACCTGGCTGGTCACAAAATTTGCATG 5835

Qy   922 tgyaaygargaryytnaaygtnaayccncargagaayggngaraayathwsntggacntgy 981
          |:||||||||||| |||: |||||||||||:|||||||:||||:|| ||||||||||::
Db   5834 AGTRAC-AGGAGCCAAATGCTAATTCCCAAGACAAATGGGGAAAAAGTCTCAGGGCATGT 5776

Qy   982 caarngqnwswsncarcaarwsnathaarwsnytng-cntgqmgnccnmgnmgnaartggtt 1040
          :||:| || ::|| ||| || :|||| :||| |||||||:|||| ||:||||| |||||
Db   5775 CAGAGGTCTTTATAGGCAACCCCTCCCATCACAGGTCCTCCAGAGGTATCAGGAAAAATGGTTA 5716

Qy   1041 ytgyggnacnggghccnggnwsnytntgtytgytgtncarccnmgnaayytngtnccntgygt 1100
          |:||||||| ||| || ||||:||||||:||||| |||:| |||:||:|||| |||||::
Db   5715 TTGTTGGCCAGGCCGCCGGGTCCTCATGCTGTGTGTCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy   1101 ncengtna------------- 1126
          ::|||| || aywsngcngtngcnwsng
Db   5715 AGTCAAGAATTGGGTTTGGAAAACCTTCCCTAGCATTAAAGAGGADATGTCGGAAAATGCCT 5476

Qy   1127 argngncnwsnccnaarccntggcaryytnccnwsnggngtngaccngtngggcnaara 1186
          |||:|||::|| || ::||| ||| |||| |||||||:|||||::||||||:||||:||
Db   5595 GGAGTGCAACCCCAAGCCTTGACAGCTTCCATGCTGTTGAGACTGGCGAGTGCACAGA 5536

Qy   1187 arwsnmgnathgargtntgggarccncncnathmgnttycaraararathtayggnaaycent 1246
          |:|||:|||||||:|||| ||||| :||||| ||||||||||:||| :| |:||||||::
Db   5535 AGTCAAGAAATTGGGTTTGGAAACCTTCCCTAGCATTAAAGAGGADATGTCGGAAAATGCCT 5476

Qy   1247 ggatgccnmgncararttygcngtngggnwsnacnwsngcnmgng 1306
          ||||:|||||| ||| ||| |||||||:|||||||| ::|||||
Db   5475 GGATCCCAGTCAGAAGTTTGCTGCAGGACCCCATGGAGATCCTCTGCCAGGG 5416

Qy   1307 tngtncararggnaaygtngggnggncaycnncacaymgncnncawsnggngnccnw 1366
          :|| ||:||||||:||||||||||||| |||||||||||:|||:||||||:|||
Db   5415 CAGTGCAGAAGGAAAATGGGGTCAGACCCCACACAGTCCCTACTGGGGCCACCAC 5356

Qy   1367 snwsnmgncngtnmgnmgnwsncncncnwsnwsmgnytncaraarggnmgnwsnacng 1426
          ::|: |||||||:||||||||||||| :||||||||| :|:|||||||:||||||||
Db   5355 CTAGTGGAGCTGTGAAGAAGAGGTCTCC------AGACCCCAGAAATGGTAGATCCACCG 5303

Qy   1427 aywsnytncarcaygtncngaraaarwsnacngayacncaartgtycarccngtnaargcng 1486
          |:|:|| ||||||||:||||:|||:||:||||||:|| ||||||:||||||||:||||:||
Db   5302 ACAGCGTTGACACCGTGTACCTGGACACTCAATGCCAGAAAAGCTCCAGACAACTGAAAAGCAG 5243

CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant

RESULT 4
AAS03690/c
ID   AAS03690 standard; DNA; 9236 BP.
XX
AC   AAS03690;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Rhesus gene locus: downstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999; 99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA, Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 10; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   downstream Rhesus box of D positives. The Rhesus genes locus
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant

Qy   1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
          :| ||:|||||||||||||| ||:||||||||:|| ||||:||||||||:||||||||||
Db   5242 CTGAGAGGGAGCGTGTACCCTGCAAAAGGTACAGGGGCAGAGCTGCCAAAGCACCATAGGGAA 5183

Qy   1547 thtayytnytncayytgycaygayytngaaytnmgncayyggngtnaarmgngaycayttyg 1606
          :| || |||:||||||| || :|||||| ||||||||:||||||||||| ||| |||||||
Db   5182 CCCACCCTTCCATCTCGTGACCTGGACCTTGAGAGTGTGAGATGTGGAGTCAAAAGGAGTCATTTTG 5123

Qy   1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggncngtnccnytnt 1666
          |||||||||||||| ||| ||||| ||:||||:|||| ||||:|||||||:|||:||::
Db   5122 GAGCTTAAGCATTTGACTGCCCCACTCAGGATTTTGACTCTCATGGGCCTGTTAGCCTCTTT 5063

Qy   1667 gyttyggncnarttytycenttyggnacngcngnctttyacnacaartgyytntaatayycayt 1726
          |::||| ||||||||:|||:|||||||:||||||:|||:|| |||: |:|| |||:|||||
Db   5062 GTTTTGACCAATTTATCCCATTTGGAATGCTGTATTTACCCACCCCATT 5003

Qy   1727 gyat 1730
          |:||
Db   5002 GTAT 4999

CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9236 BP; 2467 A; 2000 C; 2450 G; 2450 T; 0 other;

Query Match   26.1%; Score 452; DB 22; Length 9236;
Best Local Similarity 47.4%; Pred. No. 6.3e-105;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

```
Qy   744 rgargarathwsnaarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
         :|:||||:|| ||:||| ||||:||:|:|||:||:|||||||||:||:|||||
Db  6014 AGGAGAAATTTCTAAGCACCAAAACCATTCAAAGAGGTGACTTGGGTGCTTAAAGACATT 5955

Qy   804 ywsnttyathmgngargcngnarcayaarwsnwsnngaraayytncayccngayaaygtnat 863
         :|||:||| ||:|||::||:||||:||||||:|:||||:::|::||:||||||||||
Db  5954 CAGTTTTATAAGGGAAGCAGAGCATAAAACGTTCGGAAAAATTTGCAGCCTGACAAGTGAT 5895

Qy   864 haaraaraaraa--ycnnttywsngargngnaaarttyaarytngcngcnaarathgyath 921
         :|:|:||||:|| :|||||||||:|:|:|||||:||:|:|||||||:|:|||:|:|||
Db  5894 AGAAGAAGAAAAATCCCATTTTCTCGAGGAGAAAATTCAAGCTGGCTGCAGAAAATTTGCATG 5835

Qy   922 tgyaasygargayrytnaaygtnaaygccncarcargayaayggngaaraaythwsntggacntgy 981
         :|:||| |||:||:|||||:||:|||||:|||:|||||::|:||||||||:||||:||:
Db  5775 CAGAGTCTTTATGGCAACCCCTCCCATCACAGGTATCACAGGAAAAAATGGTT 5716

Qy  1041 ytgtyyggnacnggnwsnytntgytgygtncarccnmgngayytngtncntgygt 1100
         |:|| :||||:||:||:||||:||||:|||||:||:|||:|||:
Db  5775 CAGAGTCTTTATGGCAACCCCTCCCATCACAGGTATCACAGGAAAAAATGGTT 5716

Qy  1101 nccngtna--------------------aywsngcngtngcnwsng 1126
         ||:|                      ::|||:|||||||:|
Db  5834 AGTAAC-AGGAGCAAAATGCTAATTCCCAAGACAATGGGAGAAATGTTCCCAGGGCATGT 5776

Qy  1127 arggngcnwsnccnaaarccntggcarytnccnwsnggngtngarccngtngggngcnaara 1186
         :|:|||||:|:|| ||||:||:|:|||:||||:|:||||||:|||:|||||
Db  5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy  1187 arwsnmgnaathgargtgntgngaarcncncnaathmgnttycaraararathtayggnaayccnt 1246
         |::|| ||:|::|||||:|:|:|::||||:|:|:|::||||:||||||||:|||
Db  5595 GGAGTGCAAGCCCCCAAAGCCTTGACAGCTTCCATGTGGGTGTTGAGACTGCCGAGTGCACAGA 5536

Qy  1247 ggatgccnmgncaraarttcygcngtnggngnwsntggmgnacnwsngncnmgng 1306
         ||:||||||:|:|:|||||:|:||||||:|:|||:|:||||||:
Db  5535 AGTCAAGAATTGGGGTTTGGAAAACCTTCGCCTAGATTAAAGAGGATGTGGGAAATGCCT 5476

Qy  1307 tngtncaaraarggnaaygtnggntgggarccnccncaymgntgnccnwsnggngcnccnw 1366
         |::||||:|||:||||:||:||||:|:||:||:||||||:|||:|||||:
Db  5475 GGATGCCCAGTCAGAAGTTGCTCAGGAGCAGGGCCCCTACAGAGAGATCCTCTGCCAGGG 5416

Qy  1367 snwsnmgngcngtnmgngmgnwsncncncnwsnwsmgnytncaraargmgmgnwsnacng 1426
         ::|:|||||:||:||:|||:|||||:|:|:|||||||||:|||||||:||
Db  5415 CAGTCGCAAGGGAAAATGTGGGGTCAGAGACCCCACACAGTCCCTACGTGGGGGCCACCAC 5356

Qy  1427 aywsnytncarcaygtnccngaraarwsnacngayacncarttgyarccngtnaargcng 1486
         |::||||:|:||:||:||:||:||||:|:|||:||||:||:||
Db  5355 CTAGTGGAGCTGTGGAAGAGGTCCTCC------AGACCCCAGAATGGTAGATCCACCG 5303

Qy  1487 cnggnaatggarwsntnccntayaaracngtngcngarytnacnacacngtnaggna 1546
         ||||:|||||:|:||:||:|| ||:||||||||:||:|:|||:|:|:|||:||
Db  5242 CTGAGAGGGAGGCGTGTACCCTGCAAAGGTACAGGGCCAGAGCTGCCAAGACCATGGGAA 5183

Qy  1547 thtayytnytncaygtcygcygayytngagcaygyhgmgnaaymgngaycayttyg 1606
         |:|| |:||||:||::||:|||:|||:|||||:|:|||||:|||:|||:||||||
Db  5182 CCCACCCCTTCCATCTGCGTGACCTGGAGTGTGAGATGTGACTGCAAAGGAGATCATTTTG 5123

Qy  1607 gngcnytnmgnttygaytgyccnacncangnyttymgnacntayatgggncngtnccncyntt 1666
         |||:||||||:|||:|||:||:|:|:|||::||:||:||||:|:|||:||:||
Db  5122 GAGCTTTAAGATTTGACTCCCACTGGATTTGGACTGTCTCATGGGCCGTAGCCTCTTT 5063

Qy  1667 gyttyggncarttytyccnttygggnacngcngtnttyacnacaartgyygtntayytnaaytncayt 1726
         |:|||:|| |:|||:|:||||:||||:|:|||:|:|||:|:||:||::||||:|:||
Db  5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCATT 5003

Qy  1727 gyat 1730
         |:||
Db  5002 GTAT 4999
```

RESULT 5
AAS03688/c
ID AAS03688 standard; DNA; 9241 BP.
XX
AC AAS03688;
XX
DT 29-AUG-2001 (first entry)
XX
DE Rhesus gene locus: hybrid Rhesus box of RHD negatives.
XX
KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS Homo sapiens.
XX
PN WO200132702-A2.
XX
PD 10-MAY-2001.
XX
PI Flegel WA, Wagner FF;
XX
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PS Disclosure; Fig 8; 135pp; English.
XX
CC The sequence represents the coding sequence of Rhesus gene locus:
CC hybrid Rhesus box of RHD negatives. The Rhesus genes locus
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box. The RHD and RHCE genes are located at
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other;

Query Match   26.1%; Score 452; DB 22; Length 9241;
Best Local Similarity 47.4%; Pred. No. 6.3e-105;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

```
Qy   744 rgargarathwsnaarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
         :|:||||:|| ||:||| ||||:||:|:|||:||:|||||||||:|| ||:|| ||
```

```
RESULT 6
AAD05134

Db  6014  AGGAGAAATTTCTAAGCACAGCAAAGCCATTCAAGAGAGTGACTTGGGTGCTGTTAAAGACATT  5955

Qy   804  ywsnttyathmgngargcngarcayaayrwsnwsngaraayytncayccngayaaygtnat    863
Db  5954  CAGTTTATAAGGGAAGCAGAGCATAAAAATGTTCAGCCTGACAATGTGAT              5895

Qy   864  haaraaaaraa--ycanttywsngaraaryttyaaarytngcngcngaraathtgyath     921
Db  5894  AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCACGAAAATTTGCATG  5835

Qy   922  tgyaaygarggaryghaaaaygccncargagayggaaygggngaaraayayathwsntggactgy  981
Db  5834  AGTAAC-AGGAGCGAAAATGCTAATTCCCAAGACACAATGGGGGAAAAATGTTCCCAGGGCATGT  5776

Qy   982  caringnwswsncarcarwsnathaarwsnoytng-cntggmgnccnmgmgnnaartggtt   1040
Db  5775  CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCCAGGAAAAATGGTT   5716

Qy  1041  ytgyggnacnggnccnggnnwsnytntgytgygtncarccnmgmgnngayytngtncntgyyt  1100
Db  5715  TTGTTGGGCCACAGGCCCCGGGGTCCTCATGCGTGTGTCGACCTTAGGGACTTGGTCCTGCAT  5656

Qy  1101  ncongtna------                          aywsngcngtngcnwsng      1126
Db  5655  CCCACCCCAGTCAGAAGTTCCCAACCATGACGGAGGCAAAGGTAGAGCTTCGGGCGTCTGTAGCTTCGG  5596

Qy  1127  argngncnwsnccnaaarccntggcaryytnccnwsngggngtngaccngtngggngcnaara  1186
Db  5595  GGAGTCCAAGCCCCAAGCCTTGACAACTTCCATGTCGTGTTGAGACTGCACAGA         5536

Qy  1187  arwsnmgnathgargtntggaarccnccnccnwsnsnmgnyttycaraarathtaygngnaayccnt  1246
Db  5535  AGTCAAGAAATTGGGGTTTGGAAAACCTTCGCCTAGAATTAAAGAGGANTGCCGAAAATGCCT  5476

Qy  1247  ggatgcngncaaraarttygcngtnggngtnggngnnwsnwsntggmgnacnwsngcnmgmgng  1306
Db  5475  GGATGCGCAGTCAGAAGTTTTGCTGCAGGAAGCAGGGCCCTCAGTGGGAGATCCTTCTGCCAGGG  5416

Qy  1307  tngtncaaraargngnaaygtnggggntnggngnaarccnggtngggngcnaara       1366
Db  5415  CAGTGCAAGCCAAGGGAAATGTGGGGTCAGAGACCCCACACAGTCCCTACTGGGGCACCAC  5356

Qy  1367  snwsnmgnngcngtnmgnmgnnwsncccccnwsnsnmgnytncaraarggnmgnwsnacng  1426
Db  5355  CTAGTGGGAGCTGTGAGAGAGGTCCTCC------ AGACCCCAGAAATGGTNAGATCCACCG  5303

Qy  1427  aywsnytncarcaygtncngaaraarwsnacngayacncaartgycarccngtnaargcng  1486
Db  5302  ACAGCTTGCACCGTGTACCTGCAAAGGTACAGGGCGCAGAGCTGGAAAGCAG          5243

Qy  1487  cnggnatggatwsngtncontayaaracngtngtngngaryytnacnaaracngtngngna  1546
Db  5242  CTGGAGAGGGAGGCTGTTGACCCGCTGCCCAAGACCATGGGAA                   5183

Qy  1547  thtayytnytncaytgtycaygayytngaygtnaarmgngaycayttyg            1606
Db  5182  CCCACCCCTTCCATCTGCGTGACCTGGATGTGGAGTCAAAGGAGATCATTTG         5123

Qy  1607  gngcnytnmgnttygaytgygccnacngngnttymgnacntayatgggwsnacngtnccnytnt  1666
Db  5122  GAGCTTAAGATTTGACTGCGCCCACTGGAATTTGGACTCTCGTNAGCCTCTTT        5063

Qy  1667  gyttyggnacarttytyccnttyggnacncngngtnttyacncaartgyytntatytnaayt  1726
Db  5062  GTTTTGACCAATTTATCCCAHTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCATT   5003

Qy  1727  gyat  1730
Db  5002  GTAT  4999
```

CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;

Query Match   24.8%; Score 429.8; DB 22; Length 3203;
Best Local Similarity   46.5%; Pred. No. 1.3e-99;
Matches 476; Conservative 211; Mismatches 297; Indels 40; Gaps 6;

Qy   745 gargarathwsnaarcarcaarwsnathcargargtnaactngtgggtnaytnaargcntty 804
          |::||:||:||..::|:| |:|||:||| ||||:||| | ||||||||:||| |||:||
Db   1639 gaagaaattctaagcagcagaggaggcaagcgattcaaaggaggtgatttgggtgctgttaaaaagcatt 1698

Qy   805 wsnttyathmngargcngarcayaarwsnwsngarayytncaytccngayayytnath 864
          |.::|:|.|:| :::|| | ||:::||||:|| :|| |||| ||||:|||:||| ||||
Db   1699 catttaaaggyaaac--agcataaaaaatttggaaaaatttgcagccagttgatgcagca 1756

Qy   865 aaraaraaraayccnttywsngargnaarttyaaarytngcngcngarathtgyathtgy 924
          |.:|.:|||:|.::|| |:||||| :|||||||||:|||| :||| :|||:||:|||:|
Db   1757 gaagagagaaactgttttttgaggagaaatcaagctggctgcagaaatttgcataagt 1816

Qy   925 aaygaargarytnaaytnaaycnncaargayayayggngaraaayathwsnggacntgycar 984
          ||:|||| :||||:||| ||:| :||||||:||:||||| ::|||||||:||||||||
Db   1817 tacagggagctgaatgttaatccacagacaatgggaaaatgtctccaggcatgtcac 1876

Qy   985 mgnwsnwsncacaarwsnathaarwsnytngcnnngnmgnaar---tggtty 1041
          ||:.:|.|:|| ||:.:|||:|| |||| :|| :::|||||| |||:||
Db   1877 aggcttcatggcagccccctcccatcacagacccctgaggactaggagaaaatgttt 1936

Qy   1042 tgyggnacngngnccnggnwsnytntgtgytgncarccnmngnga yytngtncntgygtn 1101
          |:| |:||| || ::|| |::|||:|||||:|||| || ||||:|| |||||:|:||||
Db   1937 tgtgtgggcaggtccaagtcccatgctgtgtgcagcctaggtgctgttgtgcctgcatc 1996

Qy   1102 ccngtnaaywsngcngtngc------------------ 1129
                                          nwsngarg
Db   1997 ccagctgctccagccattccatctgggtggtgttcatggttcagagg 2056

Qy   1130 gngcnwsncnaarccntggcaryytnccnwsngggntngarccngntnggngcnaaararw 1189
          |:||::|::|| ::||||| |:|||||:|::||||::|| :||:||:|||||:|||| |
Db   2057 ccagctgctccagccatcctaaaagggctaggttgttgagcctgcaggtgcatggaagt 2116

Qy   1190 smngnathgargtngtnggarccncncnatchmgnttycaraarathtayggnaayccntgga 1249
          |:|||:|| |:||:|||||||:|| :|.|||:||||||||:||||:||:|||||||:||
Db   2117 caagaattgaggtgttgggaacctccactcagattcagaagatgtatggaatcacctgga 2176

Qy   1250 tgccnmncaaraarttygcngtnggng tnggngnwsntggmgnacnwsngcnmgngtng 1309
          ||:| :|||||:||:||| :|| |||| :|||:|:.:|:|||| :|||| :||:|||
Db   2177 tgctcaggcaaaagtttgctctccaggggccagagacctctgctaggtcag 2236

Qy   1310 tncararggnaaytgnggnggngarccnncaymgntgntccnwsnggngcnccnwsnw 1369
          :||||:||| |||| |||| |||| |||:||:||||| ::||:|||||||| ::||:|
Db   2237 tg-tgaaggaaatgtggggttggagcccacagaatccctactgggcaccacta 2295

Qy   1370 smnngngcngtnmgnwsnccnccnwsnmngnytncaraargmngnwsnacngaryw 1429
          :||:|:|| ::| |:.|.|:||:||::| |:||||:||||:||:.:||| |||
Db   2296 gtgggagctgtcgaagaggccacgtccttcagaccccagagatggtagatccaccga 2355

Qy   1430 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1488
          :.:||||||:|.:||||||||:|:|:||||:||:|||||||||||||||:||||| :|
Db   2356 gctgtaccgtgccctggaaaagcccacctaaacgccagccgtgaaagcagcca 2415

Qy   1489 ggnatggarwsngtnccntayaacngtngcngarytnacnaatacngtngngtnath 1548
          |:||:|||.:|:||:||:||||||:||:| |||:||:|||:|||:||:||.::||:|
Db   2416 ggagtgggactataccctgtgaaacagcaggagagctgcccaagctaaggaaac 2475

Qy   1549 tayytnytncaytgycaygayytngaygtngayctnmgncayggyggngtnaaarmgngaycaycttyggn 1608
          |:::|.:|| |||| |:|| |||||| :|||| |||| :|:.:||||||||:|| |:||||| |||||:||
Db   2476 tacctcttgcatccatcatgtgacctggatgtgaactcaggagtcaggagatcatttgga 2535

Qy   1609 gcnytnmgnttygayctgyccnacncngnttymgnacntaytgggncnccngt-nccnytntg 1667
          |:||:|||||:|| :|:|||:||:||:||| :|||| :|||:||||: |||:.:||:||:|||
Db   2536 acgntataattgactgcctcgctgacttttgggacttgcatggggctttgtaacgcctttg 2595

Qy   1668 yttyggncarttyttyccnttyggnacncngcntgntntyacncartgyytntayytnacaytg 1727
          :||||||||:||:.:|||:||||||| :|||||| |||:||||||:||:.:||||| |||
Db   2596 ttgtgggcagttctccccatttggaatggctgtgtattaccattacctgtgatccccaatg 2655

Qy   1728 yatg 1731
          :||
Db   2656 tatg 2659

RESULT   7
AAL03635
ID   AAL03635   standard; DNA; 6063 BP.
XX
AC   AAL03635;
XX
DT   21-NOV-2001 (first entry)
XX
PN   WO200155320-A2.
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186330.
PR   16-MAR-2000; 2000US-0189074.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
XX
OS   Homo sapiens.
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6323.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 01-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231968.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 13-OCT-2000; 2000US-0234997.
PR 13-OCT-2000; 2000US-0234998.
PR 02-OCT-2000; 2000US-0239937.
PR 02-OCT-2000; 2000US-0239935.
PR 29-SEP-2000; 2000US-0237040.
PR 29-SEP-2000; 2000US-0237039.
PR 29-SEP-2000; 2000US-0237038.
PR 29-SEP-2000; 2000US-0237037.
PR 27-SEP-2000; 2000US-0236802.
PR 26-SEP-2000; 2000US-0236370.
PR 25-SEP-2000; 2000US-0236369.
PR 25-SEP-2000; 2000US-0236368.
PR 25-SEP-2000; 2000US-0236367.
PR 20-OCT-2000; 2000US-0236327.
PR 20-OCT-2000; 2000US-0235836.
PR 20-OCT-2000; 2000US-0235834.
PR 20-OCT-2000; 2000US-0235484.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0244477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249228.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251998.
PR 05-DEC-2000; 2000US-0256719.
PR 05-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251669.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-465570/50.
XX
PS Disclosure; SEQ ID NO 6323; 1297pp + Sequence Listing; English.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 6063 BP; 1655 A; 1162 C; 1259 G; 1987 T; 0 other;
XX

```
Query Match             24.7%;  Score 427.6;  DB 22;  Length 6063;
Best Local Similarity   46.4%;  Pred. No. 8.4e-99;
Matches  475;  Conservative  211;  Mismatches  298;  Indels  40;  Gaps  6;

Qy    745 gargarathwsnaararcarwsnathcargaargtnacntgggtnytnaargngntty 804
          |:|:||:||::|:.|:|||:.:||:||:||:||.||:||||::||.:|||| ||:
Db   3481 gaagaaatttctaagcagcaagcaattcaagaggtgattttgggtgctgttaaaagcattt 3540

Qy    805 wsnttyathmngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 864
          ||:|.||:|::|||  ||:||.|.|:||:|::||..|:||||:||.||.||.|||:|
Db   3541 catttaaaagggaaac--agcataaaaatttggaaaaatttgcagccagtgatgcagca 3598

Qy    865 aaraataaraayycnttywsngarggnaarttyaarytngcngarathtgyathtgy 924
          |:|:|.||:||.|:|| | |:|||:|||:||::||.|:||:.|||:|:||||:|:|:
Db   3599 gaagagagaaacctgttctttggagagaaatccaagctggctgcagaaattgcataagt 3658

Qy    925 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntgacntycar 984
          ||:||.|:|:|.|:|:||:||:||:||:||:||.|||:||.||:||:|| |:|:|||
Db   3659 tacagggaggctgaatgttaatcccaagacaatgggggaaaatgtctccaggcatgtcac 3718

Qy    985 mgnwsnwsncatcarwsnathaaarwsnytngcntggngnccnmgnmgnaar---tggtty 1041
          :|:|:|:.:: |||:.: |||:|: |: |: ||||: ||||:||||||||| |||:||:
Db   3719 aggtcttcatggcagcccctcccatcacagacccctgaagactaggaggagaaaatggttt 3778
```

Qy　1042　tgyggnacnggngnwsnaytyntggytgyytyncarccmngngayryytngtnccntgygtn　1101
　　　　　||:||　||　||　||||　||:::||||||||　:||:||　:|||||　||:　|||：||
Db　3779　tgtgggccaggnccaggctcccacatgctgtgtgcagcctaggagacttggtgccctgcatc　3838

Qy　1102　ccngtnaaywsng----------------------------cngtngcnwsngarg　1129
　　　　　||　:||||:|　　　　　　　　　　　　　　　　:|||||||:|
Db　3839　ccagctgtcccagccattgctgctaaaaggcgcgagtacagctctgccgtgttcaagg　3898

Qy　1130　gngncnwsncnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw　1189
　　　　　:||||　||||||||||||　|||||　||:||:||||||||||||||||||||||
Db　3899　gtgcaagccccaaacttggcagcttccattggtctgtcttgagcctgcacatggaagt　3958

Qy　1190　snmgnathgaargtnggggarccnccnnathmgnttycaraaratathtaygggnaayccntgga　1249
　　　　　:||||||　:||||||||||||||||||||　||　||||||　||:|||　||　||||
Db　3959　caagaattgagggtttgggaacttccatctagattccaagaatgtatggaatcacctgga　4018

Qy　1250　tgccnmgncaraattygcngtnggngtnggngnwsnwsntcggmgnaacnwsngcnmgngtng　1309
　　　　　|　||||||||||||||||||||||　|:|||||||||||||||||||||||||
Db　4019　tgctcaggcaaaagtttgctccaggggcgagccttcatggagaacctctgctaggggcag　4078

Qy　1310　tncaraargngnaaygtnggngtngarccnccncaymgngtnccnwsnggngcncncnwsnw　1369
　　　　　　　:||||||||　:|||||||||||||||||||||||||||||　||||||||||||：
Db　4079　tg-tgaaggaaatgtgggggtttggaagcccccacacagaatcctactgggcaccacacta　4137

Qy　1370　snmggncngtnamgnmgnwsncnccnwsnwsnmgnytncaraarggnmgnwsnacngayw　1429　.
　　　　　||||||||||||　:||||||||||||||||　:::|||||||||||||:|||||
Db　4138　gtgggagctgtgcaagaaggccacgtccttcaggaacccagaatcgtagatccaccgaca　4197

Qy　1430　snytncarcacaygtnccngaraarwsnacngayancaartgycarccngtnaargcng-cn　1488
　　　　　:||||||||||||||　||:||||||||||||||||||||||||||||||||||||-||
Db　4198　gctctgtacggtgccactggtgaaaagccacaccctaaacgcccgtgaaagcagcca　4257

Qy　1489　ggnatggarwsngtnccntayaaracngtngtngcngarytnacnacaracngtnggngath　1548
　　　　　|:||||||||||　||:||||||||||||||||||||||||||||||||:||||||:||
Db　4258　ggagtggggactataccctgtgaagccacacagggcagagctgccccaagactaaggaacc　4317

Qy　1549　tayytnytnycaytgycyaytngayytngmgncaggtgtnaarmgnaaycaytygggn　1608
　　　　　||||||||||||||||::|||||||||||||||||||||||||||||:||||||||||:
Db　4318　taccctcgcatcattgtgacctggatgttgcaggagtccaggagga tcatttgga　4377

Qy　1609　gcnytnmgnttygaytggycnacngggntyymgncncntayatgggnccngt-nccnytntg　1667
　　　　　||||:||||||||||||||||||||:||||||:|||||||||||||||-|||||||:||
Db　4378　acgttataattgactgcctcgcggattttggacttggcttgtcatgggcttaacgccttg　4437

Qy　1668　yttyggncarttytycncnttyggnacngcngcntnttyancartggyytntayytncaytg　1727
　　　　　|||||||||||||||||　:||||||||||||||||||||||||||||||||||||||||
Db　4438　ttgtgggcacttctcccccatttggaatggctgttattaccccattacctgtatcccatg　4497

Qy　1728　yatg　1731
　　　　　:|||
Db　4498　tatg　4501

RESULT　8
ID　　AAS44505 standard; DNA; 22509 BP.
XX
AC　　AAS44505;
XX
DT　　18-DEC-2001　(first entry)
XX
DE　　Human LEKTI DNA clone CIF978SKB_94F21 contig 11, SPINK5 exons 1-4.
XX
KW　　dermatological; PCR primer; sequencing primer; gene therapy.
KW　　eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KW　　serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
DE　　Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
OS　　Homo sapiens.
XX
PN　　WO200164747-A1.
XX
PD　　07-SEP-2001.

XX
PF　　02-MAR-2001; 2001WO-GB00897.
XX
PR　　02-MAR-2000; 2000GB-0005098.
PR　　03-MAR-2000; 2000GB-0005229.
XX
PA　　(ISIS-) ISIS INNOVATION LTD.
XX
PI　　Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
XX
DR　　WPI; 2001-582149/65.
XX
PT　　Determining susceptibility to atopic disease or carrier status of
PT　　Netherton's syndrome in humans by identifying variants of or mutations
PT　　in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT　　inhibitor -
XX
PS　　Disclosure; Page 88-94; 123pp; English.
XX
CC　　Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC　　fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC　　SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC　　(LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC　　an atopic disease in a human subject can be detected by screening the
CC　　genome for one or more polymorphic variants of SPINK5 gene and/or
CC　　expression of a variant LEKTI protein in a tissue. Carrier status of a
CC　　subject or development of Netherton's syndrome is diagnosed by screening
CC　　for the presence of loss-of-function mutations in the SPINK5 gene. An
CC　　expression vector comprising a nucleic acid encoding a serine protease
CC　　inhibitor or its functional fragment can be used in screening for
CC　　compounds with potential pharmacological activity by determining the
CC　　serine protease activity of a protein previously identified as a ligand
CC　　of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC　　asthma, eczema and hayfever.
XX
SQ　　Sequence 22509 BP; 6725 A; 4291 C; 4284 G; 7209 T; 0 other;

Query Match　　　24.7%;　Score 427.2;　DB 22;　Length 22509;
Best Local Similarity　47.1%;　Pred. No. 3.4e-98;
Matches 472;　Conservative 207;　Mismatches 303;　Indels 21;　Gaps 7;

Qy　744　rgargaratthwsnaaarcacarcarwsnathcargargtnaachtgggtnytnytnaargcntt　803
　　　　　:||:|||||||||:||||||||::|||||||||||||||||||||:||||||||||||||||
Db　14187　agaagaaatttctaagcagcagagcacttcaagagaataactctgggtgtcttaaaggcatt　14246

Qy　804　ywsnttyathmgnaargcngarcayaarwsnwsngaraayytncayccngayaaygtnat　863
　　　　　:||||||||||||||||||||||||||||||:||||:|||||||||||||||||||||
Db　14247　cagcttttaagggaggcagaacataaaagtttgaaaattgcagcctgacagtgtgat　14306

Qy　864　haararara--araayccnttywsngarggnaarttyaarytngcngcngarathtgyath　921
　　　　　:||:||:|:::||||||||||||||:||||||:||||:|:||||||||||:||:||||||
Db　14247　aaaaaaagaaaaatccattctgaggagaaattcaagccagttgcataaatttgcata　14366

Qy　922　tgyaaygargarytnaaycccnacaragayaayyggnaarayaayathwsntggacntgy　981
　　　　　|:|||||||:||:|||:|:　||||||:|||||:||||||:||||||||||||||||
Db　14367　agtagcaaggagcctgatgttaatcccaagacaaatgg-gaaaatgtctccaagagatat　14425

Qy　982　carmgnwsnwsncarrcarwsnathaarwsnytngcntggmgnccnangnmgnaar---tgg　1038
　　　　　:|||||||||||||:||:||||||||||::|||||||||||||||||:|||||-　|||
Db　14426　cagagacctttgtgtggcaagccccctcccccattacatgccagaggtttggaggaaatgg　14485

Qy　1039　ttytggngnacnggngnwsnytntgtgtgyytgygtyncarccmngngayryytngcntgy　1098
　　　　　|:|　||||||||||||||||||||:|||||:|||:|||||||||||||||||:||
Db　14486　ttcctaggcccaggcttaggtccctctgtcgtcgtgaagacacctga　14545

Qy　1099　gtnccng-----------tnaaywsngcngtngcnwsngargngngcnwsnccnaarccntg　1148
　　　　　||||||　　　　　　　:|||||||||||||||||||||||||||||||||||||||
Db　14546　gtggctgaagggggccatggctatggcggttccagttcccgaggtgcaagcctgaagccttg　14605

Qy　1149　gcarytnccnwsnggngtnggngarccngtnggngcnaarwsnmgnathgargtntggga　1208
　　　　　||:||||||||||||||||||||||||||||||||::||||||||:||||||||||
Db　14606　acagcttccacatggctgtttgagcctggagtgcacgaagtcaagaaactgaggggttgaaa　14665

```
Qy   1209  rcnccnathmgnttycaraarathtayygnaayccntggatgcnmgncaraarttygc 1268
            ::|| ||   |  || | ::| |||   |   ||| | : | ||||| | |::|:||:|
Db  14666  acctcagctctagatctc---agatatatggaaatgccctggatgtccaggcagtttc 14722

Qy   1269  ngtnggngtngggnwswsngtggmgnacnwsnggngncnwsnwsnmgngcntmgnmgnws 1328
            | ||  |||| | |||| |   |||| |  || | :||| ||   | :| |   ||
Db  14723  tgtaggggtgaggtcctcatggagaacctctgcaggcagagaagaaatgtgga 14782

Qy   1329  ntgggarccnccncaymgntgnccnwsnggngcncnwswsnmgngcngtmngnmgnws 1388
            |::||:|| || ||| | |: | || ||:|||:|| ||| |:| :||| | |   :
Db  14783  gtaggagccccccacagagtccctactgg-gcaccacttggtggagctgttagaagaga 14841

Qy   1389  ncnccnwswsnmgngtncaraargngmgnwsnacngaywsnytncaraaygtnccnga 1448
            ||   | |:::|| ||::|| || |||||:|:| |:|| :||| || |  |||||||||
Db  14842  gcctcatcctccagaacccacaggatagcttgcaccacatatgcctga 14901

Qy   1449  raarwsnacngayacncartgycarccngtnaargcngcnggnngnatggaarwsngtnccnta 1508
            || |:: | | || :|:| :||:||||| |  | | | :|||| | ||| || |||||||||
Db  14902  gaaagccacaaacactcagtgccagcctgtgaaggagcggtaccctg 14961

Qy   1509  yaaracngtngtngcngarytnacnaaracngtngggnathtayytnytncaytgycayga 1568
            :||: |||   ||| |::|  | | :|| |::|| |||  |: |:| ||:|| ||| |||
Db  14962  caaagccacagagtgtgagttgcccaagacccatgggaaccacactcttgcatcagcatga 15021

Qy   1569  yytngaygtnmgncayggngngtnaarmgngaycayttyggngncnytnmgntttyggaytgayc 1628
            |::|  |:  |||| |  || |:||: ||:| || ::|: |  | | | ||:|:||| ||:|
Db  15022  cctggatagagacatagatcaaaatatatcattttggaacttaagatttgactgct 15081

Qy   1629  nacnggntymgnacntayatggngnccngt-ncnytntgyttyggngncarttytytyccnt 1687
            ||| || |:| |  | | ||  ||||:|:| | | |::|::| || | || ||| ||:|||
Db  15082  tcctgaatttcgacttgcatggagctgtagcccctttgtttttggccaatatcccat 15141

Qy   1688  tyggnacngcngtnttyacncaartgyytntntayytncaytgyat 1730
            :|||:| | |  |: || | | || : || ::|:|| ||::|:||
Db  15142  ttggaacagctgtgtattgacccaatggctgtaccccatgtgtat 15184
```

```
RESULT  9

AAS44506

ID   AAS44506  standard; DNA; 31529 BP.
XX
AC   AAS44506;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human LEKTI DNA clone CIT978SKB_94F21 contig 8, SPINK5 exon 5.
XX
KW   Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW   serine protease inhibitor; atopic disease; Netherton's syndrome; eczema;
KW   eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory; asthma;
KW   dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO20016747-A1.
XX
PD   07-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-GB00897.
XX
PR   02-MAR-2000; 2000GB-0005098.
PR   03-MAR-2000; 2000GB-0005229.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Hovnanian A,  Chavanas S,  Cookson W,  Moffat M,  Walley A;
XX
DR   WPI; 2001-582149/65.
XX
PT   Determining susceptibility to atopic disease or carrier status of
PT   Netherton's syndrome in humans by identifying variants of or mutations
PT   in SPINK5, a gene encoding lympho-epithelial Kazal-type related
```

```
CC   Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC   fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC   SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC   (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC   an atopic disease in a human subject can be detected by screening the
CC   genome for one or more polymorphic variants of SPINK5 gene and/or
CC   expression of a variant LEKTI protein in a tissue. Carrier status of a
CC   subject or development of Netherton's syndrome is diagnosed by screening
CC   for the presence of loss-of-function mutations in the SPINK5 gene. An
CC   expression vector comprising a nucleic acid encoding a serine protease
CC   inhibitor or its functional fragment can be used in screening for
CC   compounds with potential pharmacological activity by determining the
CC   serine protease activity of a protein previously identified as a ligand
CC   of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC   asthma, eczema and hayfever.
XX
SQ   Sequence 31529 BP; 9686 A; 5798 C; 5934 G; 10111 T; 0 other;
```

```
Query Match      24.7%;  Score 427.2;  DB 22;  Length 31529;
Best Local Similarity  47.1%;  Pred. No. 4.6e-98;
Matches  472;  Conservative 207;  Mismatches 303;  Indels 21;  Gaps  7;

Qy    744  rgargarathwsnaarcacarwsnathcargargtnacntgggtnytnytnaaargcntt 803
            :||::||::||:: ||::||| ||:|| || |:|||||:||  ::|| |: | |::||
Db  14187  agaagaaatttctaagcagcagcagagcattcaagagagataactctgggtgctcttaaaggcatt 14246

Qy    804  ywsnttyathmgngargcngarcaayarwsnwsngaraayytncayccngayaaygtnat 863
            ::| ||::|| | |:| ||| ||:|| :|| :| || |:|| |  ||:|| ||:|:|:|
Db  14247  cagctttttaaggagcagacataaaagtttggaaaattgcaagcctgacagtgtgat 14306

Qy    864  haaraara--araaaycnttywsngargngnaarttyaarytngcngcngarathtyath 921
            :||:||:| ||:||:| |  |  ||:|| |:||  || |:|| | | | | | | :||
Db  14307  aaaaaaagaaaatccatttcctgaggagagaaattcaagccagttgcataaaatttgcata 14366

Qy    922  tgyaaygargaryytnaaytgnaaccnccargayaaygggnaraaaayathwsntggaactgy 981
            :|:| : ||| | :||: :: || |:| | || |:|| ||  || |: | | |:|| | ||
Db  14367  agtagcaagagagcctgatgttaatccccaagacacatgg-gaaaatgtctccaagagatat 14425

Qy    982  carmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tgg 1038
            || : | | | | || :|  | |:||| | ||| |  || |||   |  ||   |  |||
Db  14426  cagagaacctttgtggcagccccctccccattactacgccagaggttaggaggaaaaatgg 14485

Qy   1039  ttytgyygngacnggnccnggnwsnytntgytgygtgncarccnmgngayytngtnccntgy 1098
            |:|||:||:| | || ::| ||| |:| ||:| |:| || ||   || |: |:||:|:|:||
Db  14486  tctctctaggccaggcttaggtccctctgctgtggtcagccctgcggttgaagacaactcca 14545

Qy   1099  gtnccng--------tnaaywsngcngtncnwsngarggngncwsnccnaarccntg 1148
            ||:|| |      |: :||| |||  | | || :| ||| ||:|:| |||: ||||
Db  14546  gtggctgaaggggggccatggctcttgggctatggcttccgagggtgcaagcctgaagcctg 14605

Qy   1149  gcarytnccnwsnggngtngarccngtngggcngcnaaaraarwsnmgnathgargtntgga 1208
            || : ||| |  ||| |:|| || |||  ||:| |:|||::| |  |:||  |:|| ||||
Db  14606  acagctccacatggttgttgagcctgtgagtgcacagagaactgaggttgaa 14665

Qy   1209  rcnccnathmgnttycaraarathtayygnaayccntggatgtgcnmgncaraarttygc 1268
            ::|| ||   |  || | ::| |||   |   ||| | : | :|||| | |::|:||:|
Db  14666  acctcagctctagatctc---agatatatggaaatgccctggatgtccaggcagtttc 14722

Qy   1269  ngtnggngtngggnwswsngtggmgnacnwsnggngncnwsnwsnmgngcntmgnmgnws 1328
            | ||  |||| | |||| |   |||| |  || | :||| ||   | :| |   ||
Db  14723  tgtaggggtgaggtcctcatggagaacctctgcaggcagagaagaaatgtgga 14782

Qy   1329  ntgggarccnccncaymgntgnccnwsnggngcncnwswsnmgngcngtmngnmgnws 1388
            |::||:|| || ||| | |: | || ||:|||:|| ||| |:| :||| | |   :
Db  14783  gtaggagccccccacagagtccctactgg-gcaccacttggtggagctgttagaagaga 14841

Qy   1389  ncnccnwswsnmgngtncaraargngmgnwsnacngaywsnytncaraaygtnccnga 1448
            ||   | |:::|| ||::|| || |||||:|:| |:|| :||| || |  ||
```

```
Db  14842  gcttcatccctccagaacccagatagcacaggtagcttgcaccaatgcctga 14901
Qy  1449   raarwsnacngaycncnartgycarccngtnaargncngcnggnatggarwsngtnccnta 1508
           |::|::|| :||:|  :|::||:::||||||: :||||:  :|||: :||| :||||
Db  14902  gaaagccacaaacctcagtgccaacgcctgtgaagcagctgggaaggaggcggtaccctg 14961
Qy  1509   yaaracngtngcngarytynacnaaracngtnggnathtayytnytncaytgycayga 1568
           |::|: :| |:| | :| |:|||:|  :||:||  |:||| ||:|||:||||:||||
Db  14962  caaagccacagaggtgaagttgccaagaccacccctgtcatacagcatga 15021
Qy  1569   yytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgycc 1628
           ::||::|::| |  :|||||| | :|||||||: :::|||:||||:||||||
Db  15022  cctggatatgagacatagatcaaaatatcattttggaacttaagatttgactgcct 15081
Qy  1629   nacnggnttymgncntayatgggnccngt-nccnytntgyttyggncarttytttyccnt 1687
           :|| :|| :| ||| :||  |||| |||| :||: ||:| ||:: | :||||| |:
Db  15082  tcctgaatttcgacttgcatggcatctgtagccccttgttcttggccaatatccccat 15141
Qy  1688   tyggnacncngcngtnttyacncartgyytntayytncaytgyat 1730
           ::||||| ||| ||  :|| |||: ||::|| |:|  |::||||
Db  15142  ttggaacagctgtattgacccaatggctgtacccccaatgtat 15184
```

```
RESULT  10
AAL03634
ID   AAL03634 standard; DNA; 6063 BP.
XX
AC   AAL03634;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6322.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
```

PR	31-JAN-2000;	2000US-0179065.		PR	14-AUG-2000;	2000US-0225759.
PR	04-FEB-2000;	2000US-0180628.		PR	18-AUG-2000;	2000US-0226279.
PR	24-FEB-2000;	2000US-0184664.		PR	22-AUG-2000;	2000US-0226681.
PR	02-MAR-2000;	2000US-0186350.		PR	22-AUG-2000;	2000US-0226868.
PR	16-MAR-2000;	2000US-0189874.		PR	22-AUG-2000;	2000US-0227182.
PR	17-MAR-2000;	2000US-0190076.		PR	23-AUG-2000;	2000US-0227009.
PR	18-APR-2000;	2000US-0198123.		PR	30-AUG-2000;	2000US-0229287.
PR	19-MAY-2000;	2000US-0205515.		PR	01-SEP-2000;	2000US-0229924.
PR	07-JUN-2000;	2000US-0209467.		PR	01-SEP-2000;	2000US-0229343.
PR	28-JUN-2000;	2000US-0214886.		PR	01-SEP-2000;	2000US-0229344.
PR	30-JUN-2000;	2000US-0215135.		PR	01-SEP-2000;	2000US-0229345.
PR	07-JUL-2000;	2000US-0216647.		PR	05-SEP-2000;	2000US-0229509.
PR	07-JUL-2000;	2000US-0216880.		PR	05-SEP-2000;	2000US-0229513.
PR	11-JUL-2000;	2000US-0217496.		PR	06-SEP-2000;	2000US-0230437.
PR	14-JUL-2000;	2000US-0218290.		PR	06-SEP-2000;	2000US-0230438.
PR	26-JUL-2000;	2000US-0220963.		PR	08-SEP-2000;	2000US-0231242.
PR	26-JUL-2000;	2000US-0220964.		PR	08-SEP-2000;	2000US-0231243.
PR	14-AUG-2000;	2000US-0224518.		PR	08-SEP-2000;	2000US-0231244.
PR	14-AUG-2000;	2000US-0224519.		PR	08-SEP-2000;	2000US-0231413.
PR	16-AUG-2000;	2000US-0225213.		PR	08-SEP-2000;	2000US-0231414.
PR	14-AUG-2000;	2000US-0225214.		PR	08-SEP-2000;	2000US-0232080.
PR	14-AUG-2000;	2000US-0225266.		PR	08-SEP-2000;	2000US-0232081.
PR	14-AUG-2000;	2000US-0225267.		PR	12-SEP-2000;	2000US-0231968.
PR	14-AUG-2000;	2000US-0225268.		PR	14-SEP-2000;	2000US-0232397.
PR	14-AUG-2000;	2000US-0225270.		PR	14-SEP-2000;	2000US-0232398.
PR	07-JUL-2000;	2000US-0225135.		PR	14-SEP-2000;	2000US-0232399.
PR	14-AUG-2000;	2000US-0225447.		PR	14-SEP-2000;	2000US-0232400.
PR	14-AUG-2000;	2000US-0225757.		PR	14-SEP-2000;	2000US-0232401.
PR	14-AUG-2000;	2000US-0225758.		PR	14-SEP-2000;	2000US-0233063.

PR	14-SEP-2000;	2000US-0233064.
PR	14-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6322; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 6063 BP; 1656 A; 1163 C; 1262 G; 1982 T; 0 other;

Query Match   24.6%; Score 426.4; DB 22; Length 6063;
Best Local Similarity   46.4%; Pred. No. 1.7e-98;
Matches 475; Conservative 210; Mismatches 299; Indels   40; Gaps   6;

Qy   745 gargararathwsnaarcararwsnaathcargargtnacntgggtnytnytnaargcntty 804
            |:||:||:||::|| |||||:||:::|||:|:|||:|||::||:|| |:|||| ||::|
Db   3481 gaagaaattctaagcacagcaaagcattcaagagagtgattgggtgctgttaaaagcatt 3540

Qy   805 wsnttyathmgngargcngarcayaarwsnwsngaraayytncaycvcngayayaaygtnath 864
            ||:| ::| |||:| :||||||:|| ::|||:| |:|| :||| ||||:|
Db   3541 catttaaaagggaaac--agcataaaaatttgaaaatttgcagccagttgatgtgcagca 3598

Qy   865 aaraaraaraayccnttywsngargngararathtgyathtgy 924
            ||: |:| ||:|| |||| |::|| ||| |::|
Db   3599 gaagagagaaaactgttttttgaggagaaattcaagctgcggctgcagaaatttgcataagt 3658

Qy   925 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggactngycar 984
            |: :||:||||:|| |||:|| |||:|| |:::||||||||||:||| |::: ||||:||

Db   3659 tacaggagagctgaatgttaatctccaagacaatgggagaaaatgtctccaggggcatgtcac 3718
Qy   985 mgnwsnwsncarcarwsnathaarwsnytncntgggmgnccnmgnmgnnaar---tggtty 1041
Db   3719 aggtcttcatggcagcccccctcccatcacacagacccgaagactaggaggaaaaatggttt 3778
Qy   1042 tgyggnaacnggnccnggnwsnytntcgytgyygtncarccnmgnaayytntgytnccntgyygtn 1101
Db   3779 tgtgtgggccaggcagcccaggtcccatctgtcgtgcagcctaggttggtccctgcatc 3838
Qy   1102 ccngtnaaywsng---------------------------cngtngcnwsngarg 1129
Db   3839 ccagctgcttccagccattgctaaaagggccgaggtacagctctgccctggttcagagg 3898
Qy   1130 gngcnwsnccnaaarccnntgycaarytnccnwsnggngtngarccncngtnggngcnaaraarw 1189
Db   3899 gtgcaagccccaaaacttgcagacttccatgtgttgaggctgcaggtgtcatggaagt 3958
Qy   1190 snmgnnathgargtntgggarccnccnacathmgnttycaraaraathtaygngnaayccntgga 1249
Db   3959 caagaattgaggttgggaacctccatctagaattcagaaagatgtatggaatcacctgga 4018
Qy   1250 tgccnmgncaraarttycgncgtngngtnggngwsnwsngacnmgnwsngcnmgngtng 1309
Db   4019 tgctcaggcaaaagtttgctccaagggcaagaacctctgcatggcag 4078
Qy   1310 tncaraargnaaygtnggngtnggngtnggngtnccnwsnggngcncnwsnw 1369
Db   4079 tg-tgaaggaaatgtggggttgggaggtggagcccccacagaatccctactgggcccacta 4137
Qy   1370 snmgncngtnmgnmmgnwsnccncnwsnwsnmgnytncaraaggnmgnwsnacngayw 1429
Db   4138 gtggagctgtgcaaagaaggccacgtgtcctcagaccccagaatggtagatccaccgaca 4197
Qy   1430 snytncarcaygtncngaraarwsnacngayacncaartgycarccngtnaargcng-cn 1488
Db   4198 gcttgtaccgtgcacctggaaaagccaagtccaaacgccccgtgaaagcagca 4257
Qy   1489 ggnatggaarwsngtnccntayaaracngtngtngarytnacnacaaarangtngnath 1548
Db   4258 ggagtgggggactataccctgtaaagccacaagggggcagcccaagactacaaagggaac 4317
Qy   1549 tayytnytncaytgyacayayaytngayytnmgncaygyggngtnaarmgnaycaycaytyggn 1608
Db   4318 tacctcttgcatcatgtgacctggatgtgaagacatggagtcaagagagatcaatttgga 4377
Qy   1609 gcnytnmgnttygaytgycnacangngnttymgnacntayatgggngccngt-ncnnytntg 1667
Db   4378 acgttataatgactgcctgcctggatttggacttgcatgggcttgtaacgcctttg 4437
Qy   1668 yttyggncaartyytyccnttyggnacngcngtntyacncaartgyytntayytncaytg 1727
Db   4438 ttgtgggcacttctccccattggaatggcgtgtatttaccctgctattcccatttg 4497
Qy   1728 yatg 1731
Db   4498 tatg 4501

RESULT 11
AAH13678
ID AAH13678 standard; cDNA; 1736 BP.
XX
AC AAH13678;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10539.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX

```
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-0248036.
PR   27-AUG-1999; 99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0837767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 10539; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesizing 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide comprising a sequence
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;
```

Query Match   24.3%; Score 421.4; DB 22; Length 1736;
Best Local Similarity   46.2%; Pred. No. 1.1e-97;
Matches 474; Conservative 211; Mismatches 293; Indels 47; Gaps 6;

```
Qy    744 rgargarathwsnaarcarcarwsnathaargargtnacntgggtnytnaargcnt  803
Db    110 agaagaaatttctaagcagcaaagcatccaagaggtgattggatacgtaaagcatt  169

Qy    804 ywsnttyathmngnqargcngarcaayaarwsnwsngaraayytncaycngayaaygtnat  863
Db    170 tagtttataagggaagcagcataaaagtttggaaaatttgcagcattactatgccgat  229

Qy    864 haaraaraaraayccnttywsnqargnaarttyaarytngcncngarathtgyathg  923
Db    230 agacaagagaaaaccanccatttctcggggagagaattcaaccaagctcgcgaaa  289

Qy    924 yaaygargarytnaaygtncacargayaayggnqgaraayathwsntggacntgyca  983
Db    290 tacaagagcctaatgttagtccccagaccactgggaagaagtctccagaccatgtca  349

Qy    984 rmgnwswsncarcaarwsnathaarwsnytngcntggmgnccnmgnaartggtttytg 1043
Db    350 gagacttcaccaagccctcctatcacaggcccaga-----aggaaaagtggtttg  404

Qy   1044 yggnacngngnccngnwsnytntgtgygtncarccnmngnayytngtnccntgyytncc 1103
Db    405 tggcctggccccaggtcccttgctgtgtgcaacctagggatgtggcaacctgtgtcc  464

Qy   1104 ngtnaayws----------------------ngcngtngcnwsnga 1127
Db    465 agctgctgcttcagctgtggctggaaagggccaatgtacagctcaggttgtggcctcaga  524

Qy   1128 rggngcnwsnccnaarccnttggcarytnccnwsnggngtngarccngtnggngcnaaraa 1187
Db    525 gggtggaagcccaagcttggcagctgtccacatggttgttgagcctgcaggtgcaacaga  584

Qy   1188 rws-nmgnathgargtntggargarccnccnaathmgntycaaraarathtayggnaayccnt 1246
Db    585 gtcaagaaattgaggtttgggaacctcctactagattc---agatgtatggaaatgtct  641

Qy   1247 ggatgccnmgncaraarttygcngtnggngtnggngnwsnwsntggngnacnwsngcnmgng 1306
Db    642 agatgcccaggcaaaagtttgctat-ggggtgggccccatagagaacctctctctaggg  700

Qy   1307 tngtncararggnaaygtnggngtgggaarcncnccncaaymgngtnggntncncnwsnggngtncccnw 1366
Db    701 cagtgtggaggaaatgtggggnttggagcccccacacagagtccctactgggggcactgc  760

Qy   1367 snwsnmgngncngtnmgnmgnwsncncccnwswsnwsmgnytncaraaargngmgnwsnacng 1426
Db    761 ctagtagagctgtgaagaggggccattgtcttctcagaacccctgaatggcagaccactg  820

Qy   1427 aywsnytncarcaygtncngaraarwsnwsnacngayacncarttgycarccngtnaargcng 1486
Db    821 acagcttgcaccatgtgcctggaaaagctgcagacactcaacgcccgtgaaagcat  880

Qy   1487 cnggnatggarwsngncnccntayaaracngtngtngcngarytnacnaaracacngtnggna 1546
Db    881 ccggagggaggctgtaccctgccaagccacaggggcagagtcaccaagaccatggaa  940

Qy   1547 thtayytnytncaytcgycaygayytngaygtnmgncayggngtnaaarmgnmgnqaycaytt yg 1606
Db    941 ctcacctcttcgcatcagtgtgacctggattctgagacctggaatcacattttg 1000

Qy   1607 gngcnytnmngnttygaytgyccnacngngnttymgnacntayatgggngnnccngt-nccnytn 1665
Db   1001 gagcttaacatttgactgccccactgaattctgacttgcatgggacccgtaacccat 1060

Qy   1666 tgyttycggncaarttyttycnttygnacngacngntttyacncatcgyytntnaayytncay 1725
Db   1061 tgtttcggtccaattctctcattggaatggctgtatttaccccaatacctgcacacccat 1120

Qy   1726 tgyat 1730
Db   1121 tgtat 1125
```

RESULT 12
```
ID   ABA90193 standard; DNA; 465237 BP.
XX
AC   ABA90193;
XX
DT   11-FEB-2002 (first entry)
XX
DE   Human oestrogen receptor alpha gene.
XX
KW   Human; oestrogen receptor alpha; ESR-alpha; ER; chromosome 6; Syne-2;
KW   synaptic nuclei expressed gene 2; haplotype; cytostatic; osteopathic;
KW   cardiant; vasotropic; gene therapy; vaccine; cancer; osteoporosis;
KW   cardiovascular disease; oestrogen receptor; ds.
XX
OS   Homo sapiens.
XX
```

```
PN   WO20016229-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05358.
XX
PR   22-FEB-2000; 2000US-183756P.
PR   20-OCT-2000; 2000US-0692414.
PR   24-JAN-2001; 2001US-0768184.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
XX
DR   WPI; 2002-041152/05.
DR   P-PSDB; AAG68251.
XX
PT   Novel variant of estrogen receptor alpha polypeptide useful for
PT   determining the biological activity of a protein for high throughput
PT   screening and for raising antibodies that elicit an immune response in
PT   host -
XX
PS   Example; Fig 1 page 1-93; 333pp; English.
XX
CC   The present invention describes an isolated peptide (I) consisting of an
CC   amino acid sequence selected from: (a) the amino acid sequence of a
CC   variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251;
CC   or (b) a fragment comprising at least 10 contiguous amino acids of the
CC   protein in AAG68251. (I) has cytostatic, osteopathic, cardiant and
CC   vasotropic activities, and can be used in gene therapy and vaccine
CC   production. (I) is useful for identifying an agent that binds to (I), by
CC   contacting (I) with an agent and assaying the contacted mixture to
CC   determine whether a complex is formed with the agent bound to the
CC   peptide. A polynucleotide (II), encoding (I), is useful in the
CC   development of diagnostics and therapies for diseases and disorders
CC   mediated/modulated by an oestrogen receptor (ER). (II) is also useful in
CC   gene therapy for treating cancer, osteoporosis and cardiovascular
CC   diseases. The human ESR-alpha gene is located on chromosome 6. The
CC   present sequence represents the human ESR-alpha gene, which is given in
CC   the exemplification of the present invention.
XX
SQ   Sequence 465237 BP; 133988 A; 89578 C; 93946 G; 147721 T; 4 other;

Query Match             24.3%;  Score 421;  DB 24;  Length 465237;
Best Local Similarity   46.5%;  Pred. No. 2e-95;
Matches    477;  Conservative  213;  Mismatches 292;  Indels  43;  Gaps   7;

Qy     746 argarathwsnaarcarcarwsnathcargaargtnacntggngtvtnytnaargngnttyw 805
           :|||:|| || ||:| | : |||||:||| ||::|| |:||||||| ||| |::
Db  438788 aagaaaattcctaagcaacaaacattcaaaagtggactgggtgttgttgagccttgaagaat 438847

Qy     806 snttyathmgnargcngarcaydarwsnwsngaraayytncayccngayaayaaygtnatha 865
           ||:|:|| | ||: ||: || :|| ||||||||||| ||:|||:||| ||||:||| ::|
Db  438848 gtttaaaagaagaaac--agcataaaagtcagaaaattcagcgcctgatgcagtag 438905

Qy     866 arararaaraaycc-nttywsngargngnarathtgyathtgy 924
           ||:||||||||||| |||:||| ||| ||| |||||||| |||:
Db  438906 gaaagaaaaaccattttttttggaggagaaattcaagctggctggcagaaaattgcataagt 438965

Qy     925 aaygargaryytnaaytnaaycncnarcagayavaaygyngaraaayathwsntggacntgycar 984
           ||: ||:||| ||||| ||||:||:|| |||||| ||::|| |||||||||:|||:||||:|||
Db  438966 aacaagagagccaaatgttaatcccaagacaatgggaaatgtctccagacatgtcat 439025

Qy     985 mgnwswsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggtt 1040
           :| |:|||| | ||:|||||:| :||||||| ||::|| || || ||||||| |||||:||
Db  439026 aggtcttcatggcagcccctcccatcacagaccggaagcctaggagaggaaaaacagtt 439085

Qy    1041 ytgyggnacnggngnccngngnwsnytntgtygyytgcayggtgcaggaggaagnsnnytnca 1100
           :||:|||| |||:|| || ::: ||||:|| ::||| |||||| || ||:||| |:||:
Db  439086 ttgtggggccagtccacaggtcccatgctgtgcagcctaggaacttggtcgcctgcat 439145

Qy    1101 nccngtnaaywsngcngtngc---------------------nwsngargngncnws 1136
           |:||| | ||||||| | |||                      ||:||:||||||||:
```

```
Db  439146 ctcagctgctccagctattgctaaaagggggctgaggtaccacggttcagaggttgcaag 439205
           | || | ||| :| || ||                   :::|| |||| || |::
Qy    1137 nccnaarccntggcarytnccnwsnggngtngarccngntngggngncnaaraarwsnmgnat 1196
           ||| || | |||: || |::  || |||:| |||: |||:: :|| ||||||:
Db  439206 cccaaaacttggcagcttcatgttggtgttgagccttgagttacacagagaagttaagaat 439265
           | |: ||||||| | ||: || |||: || | ||:: |::|| ||||:|||||
Qy    1197 hgargtntgggarccncnccnathmgnttycararaarathtayggnaayccntggatgccnmg 1256
           ||:|||||||||||||| :|| ||||| ||:| ||||||:| ||| ||| | |||||| ||
Db  439266 tgaggtttgggaacctcacttatatttcagaagatacgtggaaatgcctggatcccag 439325
           :| |||||||| ||:|: || | || |||:||| | ::| |||||:|||:|:|||
Qy    1257 ncaraaarttygcngtngnggtngggnwswsntgggmgnacnwsngcnmgntgntgncaraa 1316
           ||| || |||:|| |||:|| |:||| ::||||||:| |:||| | ||| || ||::
Db  439326 gcaaacacattcgctgcagaggtgggggccctcatggaggcctctgctaggggcaatgaggaa 439385
           ||| ||| |||| || |||| |||||| ||| ||:||| || ||| ||||| ||||||||
Qy    1317 rggnaaygtnaggntgggarccncncncancaymgngtnccnwsnggngcnccnwsnwsnmgngc 1376
           ||| ||||| ||: || || :|| | | ||||: ||||:||| || | || ::: ::||:|
Db  439386 gggaaatgtggggttggaaccccacaagtcccactggggcactgcctagtggagc 439445
           ||| || ||||| || || |||| | ||:||| ||| || | | |||:|| |
Qy    1377 ngtnmgnmgnwsnccncnccnwsnsnmgnytncaraargngngtnaarmgncaygaywsnytnca 1436
           :|| |:| | ||| | | |||: :| ||||||:| | || || ||:|: || || ||| |
Db  439446 tgtaagaggaggaccactgtcctccagacgccagaatagtagatccactgacagcttgca 439505
           :|| ||||||| |||||| :|:| ||| |: |||| |||| | ||||| |||| ||:|
Qy    1437 rcaaytnccngararaarwsnacngayacncarttgycarccngtnaargngncnggna---- 1492
           |||: || ||| ||  ||||| || || || ||| ||: | ||:||| | |:|||
Db  439506 ccatgtgcctggaaaagccacagacacctcaacgcctgtgaaagcagtcaggggtgg 439565
           |||| |:|| ||||| | ||||| ||| | ||| ||| ||| | | ||| |||||
Qy    1493 ------tggarwsngtnccntayaracngtngtngcngarytnacnaaracacngtnggna 1546
           |:| || || | | || |: | ::| || | |::|| || | | || ||
Db  439566 aggtggtgctataccctataaagcccaagagctgccaagactattgggaa 439625
           || || | ||| | | ||| | :|||||| || || || ||| ||
Qy    1547 thtayytnytncaytgycaygaytngaygtnmgncaygngngtnaarmgncaytyg 1606
           ||:|:|| ||:|| || ||:|| : ||||: |:|| |||| | || |||:||
Db  439626 cctacctcttgcatcaagcatggatgtgagacattcagtcaaaggat-atttg 439684
           ||| ||| || ||| | ||||| :|||:| |:|| |||| |||:|| ||||:
Qy    1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatggnccngt-nccnytn 1665
           |:| || | || |:| | |:||| | || |:| || ||  ||||| |:|| |
Db  439685 aagctttagaatttgactgctggtggatttagacttgtgtgggccctgtaaccctt 439744
           ||| |||| ||||| | | ||| ||||| | ||| || ||| ||| ||| ||
Qy    1666 tgyttyggncaarttyytycnttyggnacngcngtnttyacncartgyyytntayytncay 1725
           ||:||||| ||| || |:| |||||| || ||||| || |||||| | |:||| ||||
Db  439745 tgttttgggccaattcctcccatttggagcgctgtcattaccaacatgttaccccgccat 439804
           || ||||| ||| || | ||||||| | | |||:| | ||||| || |: || ||
Qy    1726 tgyat 1730
           ||:||
Db  439805 tgtat 439809
```

```
RESULT  13
AAS91984
ID   AAS91984 standard; cDNA; 1485 BP.
XX
AC   AAS91984;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27788.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
```

```
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27797.
XX
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID NO 27788; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1485 BP; 442 A; 333 C; 373 G; 337 T; 0 other;
XX
```

```
Query Match             23.6%;  Score 409;  DB 23;  Length 1485;
Best Local Similarity   46.5%;  Pred. No. 1.5e-94;
Matches  458;  Conservative 199;  Mismatches 287;  Indels  40;  Gaps  6;

Qy        787 gtnytnytnaaargnttywsnttyathmngnargncngarcayayaarwsnsngaraayytn 846
              |:||:|| ||||   | ||| ||| |  || ||| | |||:|||  |||:| |||||||| |
Db        103 gcttcttgtatgcattcagttttataaggaagcagaactaaaaatttggaaaattg 162

Qy        847 cayccngayaaygtnathaaraaraaraaycccnttywsngarggnaarttyaarytngcn 906
              |:|| || ||| |||| |||  |||    ||||||||:| ||| ||||| ||| ||||| |
Db        163 cagcctgactatgcaaatagaaaagaaaaacccatttctcggggaggaaactcaagcctat 222

Qy        907 gcngarathtgyathtgyaaygargaryaaaygtnaaygtnaaygaayggnarggngaraay 966
              ||:|| | || || ||| ||| |||| |   |  || |||:|  ||||||||  |  ||||
Db        223 gcagaaattgcctaagtagcaaggagccttaatcccaagacatcgggaaat 282

Qy        967 athwsntggacntgycarngnwswsncarcarwsnathaarwsnytngcntggmgnccn 1026
              |:||:  |||| ||||   |  ||| ||||  |  |:|| |||||||| ||:|| | |
Db        283 gtctacaggccatgcacggaacttcatggctgccccccatcacaagcctgcaggcc 342

Qy       1027 mgnmg----naartggttytgygnacngnccngngnwsnytntgytgygtncarccnmgn 1083
              |:|      |||||| |||||||| | |||||:||||| |||| | |:|||| ||| |
Db        343 aggagtaaaaagtggttcatggccagggccagggttcctgtgccgtgtgcacacagg 402

Qy       1084 gayytngtnccntcngtytnccng--------------tnaay 1110
              ||:|| |||   ||| |||  |||              :||:
Db        403 gacttggtgccctgtgtccagctgctccagctagggggcaaggtgtagaggt 462

Qy       1111 wsngcngtngcnwsnargggngcnwsncnaarccntggcaarytncnnwsngggntngar 1170
              ::|||:||||:| ||||:| || |||| | |:||||:|| | ||||||:| ||| ||||
Db        463 tgggctgtgggcttcagaggtgcaagaccaagcctgcagcttccacgtggattgag 522

Qy       1171 ccngtnggngcnaaraswsnmgnathgaargtntgggarccnccnoathmgnttycaraar 1230
              |:||:| ||| | ||  |:| ||  ||:| |:||||||||:|::|| ||:|||||||:
Db        523 cctgtgggctgcacagaagtcaagaactgaggtttgggaacctgccctggattcaagag 582
```

```
Qy       1231 athtayggnaayccntggatngcccnmgncacaraarttycgnctngtnggngtnggngnwswsnsntgg 1290
              |:|||:|||||| ||||| ||| || |  ||||||   || |:||  |:|| |||| :||
Db        583 atgtatggaaacgcctggatgcccagcaaaagtttgctcaggtgtgggaccctcatgg 642

Qy       1291 mgnacnwsnsgncnmgngtngtncaraargngnaaygtnggntnggar-ccnccncaymgngt 1349
              |:|| | || |||:||| ||| | | ||||  | |||| |:|   |   ||:|||||| |
Db        643 agaacctctgctaggcgagtgcagaggaaaatgtgggtcagagccccccacacagagt 702

Qy       1350 ncccnwsnggngcnccnwswsnmgngngcngtnmngnmgnwsncncncnwsnsmgnytnca 1409
              |:|    ||||| |||||:|:|| ||:| |||  ||  |:|| | |  ||:| |||||:
Db        703 ccctactgggacaccacctagtcggagctgtgaaggccaactgtgccaccca 762

Qy       1410 raarggngmgnwsna-cngaywsnsnytncarcayytnccngarraarwsnaacngayacart 1468
              ||:||||| | | |  ||| ||:|:||||||||||||||:|||  |::| | |||| ||:|
Db        763 gaatggtagatctaccaacagcttgcacctgtgcacctgaaaagccacagaaactcaat 822

Qy       1469 gycarccngtnaargncngngnatggarw-sngtnccntayaaaracngtngtngcngar 1527
              |:||  |:||:||  ||||:| |:|||| |  |||| | |||  | ||| || |||||
Db        823 gccagaccgtgaaagcaactaggagagggtgctgtgtaccctgcaaagcacaggtggag 882

Qy       1528 ytnacnaacarngtngnathcayyttnytnaygtnayggtytmgncayggn 1587
              |:|| ||||  |||  ||| |||||  |:|| ||||:|| ||:||| ||||
Db        883 ctgccaaggccatggggagcccacgtcttgcatcagcatgacctggatggttagacatgga 942

Qy       1588 gtnaarmgngaycayttygggngncnytnmgnttygaytgycnacngnptttymgnacntay 1647
              ||||| |:|||| ||||||| |:|||  | |||||:| | | ||||| | |:||| ||:||
Db        943 gtcaaggagatcattttgggcttaaaatttgactgccctactggattcagactgc 1002

Qy       1648 atgggnccngt-nccnytntgtytyggncaartteytycnnttyggnacncngcngntnttyac 1706
              ||||| |:||  |||    | |||||| ||::| ||||   | |||| | | |:|| ||||
Db       1003 gtgggccctgtaaccactttgtttttggccaatatctcccagttggaatggctgtatttac 1062

Qy       1707 ncartgyttnayytncaytgyat 1730
              ||:| ||||||||:||||| |||
Db       1063 tcaataactgtacctccactgcat 1086
```

```
RESULT  14
AAD05178
ID   AAD05178 standard; cDNA; 1278 BP.
XX
AC   AAD05178;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:68.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             224..268
FT                   /tag= a
FT                   /product= "Human secreted protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200134769-A2.
XX
PD   17-MAY-2001.
XX
```

PF 01-NOV-2000; 2000WO-US30040.
XX
PR 05-NOV-1999; 99US-0163580.
PR 30-JUN-2000; 2000US-0215130.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;
XX
DR WPI; 2001-308781/32.
DR P-PSDB; AAE01289.
XX
PT New isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 442; 519pp; English.
XX
CC AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 24 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1278 BP; 344 A; 264 C; 358 G; 299 T; 13 other;

Query Match   23.6%; Score 408.8; DB 22; Length 1278;
Best Local Similarity   45.8%; Pred. No. 1.5e-94;
Matches 458; Conservative 205; Mismatches 298; Indels 40; Gaps 6;

Qy   745 gargarathwsnaarcarcarwsnathcarngargtnacntgggtnytnaargcntty 804
          || | :: :|| | ||| :: | ||||:|| | ||||| :| | | ||:||
Db   280 gaagaaattctaagcacagcaaaagcattcaaagaggtgattgggtgctgttaaaagcattt 339

Qy   805 wsnttyathngargcngargcayaarwsnwsngaraaaytyncaycgayaaygtgnath 864
          ||:|| | ||| | ::| | |::| :|| |: | || |:|| | | || |
Db   340 catttaaaagggaaac--agcataaaaaaatttggaaaaatttgcagccagttgatgcagca 397

Qy   865 aaraaraaaaycncttywsnargarggnaaayathwsntggacntcgar 924
          | |:| ||||:|| | ||||| || || |:| || | ||| |||||
Db   398 gaagaagagaaacctgtttttgaggagaaatcaagctggctgcagaaattgcataagt 457

Qy   925 aaygargaytnaaygtnaaycncncargayayavngarathtgyathtgy 984
          | || | | | | | |: | || || |::|| | | ||:|||| | ||
Db   458 tacaggggagctgaatgttaatctccaagagacaaatgtctccaggggcatgtcac 517

Qy   985 mgnwswsncarcarwsnathaarwsnytngcntggmgncncmgngmgnaar---tggtty 1041
          :| : | :|| :| | :: | | |::||| | | | | | |:| |||:|||
Db   518 aggtcttcatggcagccccctccccatcacagaccctgaagaagactaggaggaaaaatggtt 577

Qy   1042 tgyggnacnggganccngnswsnytntgytygtgtncarccnmgngayytngtnccntgygtn 1101
          :|:|| | | || :: | || | | | | ||| | | | |:| || |:||
Db   578 tgtgggccaggccaggtccccatgctgtgtcagcctaggggacttggcccctgcatc 637

Qy   1102 ccngtnaaaywsngcngtngc---------------------nwsngarg 1129
          | || ||:|| || | | ::|| |:|
Db   638 ccagctgcttccagccattgctaaaaggaccgaggtacaagctctgctggttcagagg 697

Qy   1130 gngcnwsnccnaaarccntggcaarytnccnwsngggntngarccntngngcnaaaraarw 1189
          :|| | | | | || |::| | |::|| || ||:| |:: | | :| || | :
Db   698 gtgcaagccccaaacccttggcagcttccacatgtggttggcctcgcagtggcatggaagt 757

Qy   1190 snmgnnathgargtntgggarccnncnaacthmgnttycaraaraththaygngaayccntgga 1249
          |: || |: || | || | |: |::|| :|| | | :|| | || ||||| ||
Db   758 caagaattgaggttcgggaacctccatctagatttcagaagatgtattggaatcacctgga 817

Qy   1250 tgccmgncnacaaartycgncngtngggtnggnwsnwsntggmgnaacwsngcnmgngtng 1309
          :|:|:||| || ||:|:|| ||:||| | | || |: |||:| | || | ||:||
Db   818 tgctcaggcagaaaagtttgctccaggggcagagcctcatggagaacctctgctgggcag 877

Qy   1310 tncaraargnaaaytngngttgggarccncncacaymgnttnccnwsnggngcnccnwsnw 1369
          :: ||| ||:| || |||||| | |:| | | :| || || ||:|| ||:|| || |
Db   878 tg-tgaaggaaatgtggggttggagcccycacacagaatccctactgggtggcaccacta 936

Qy   1370 snmgngcngtnmgnmgnwsnccncncwsnwsnsmgnytncaraarggngmgnwsnacngayw 1429
          |: ||||||:|::|::|| | |: | |::| : :|| | | |:| | : | | ||| |
Db   937 gtggagctgtgcgaagaagccaccgtcctcaagaccccagaatggtgtrgatccaccgaca 996

Qy   1430 snytncarcaygtncngararaarwsnacngayacncaartgycarccngtnaargcng-cn 1488
          |:: || | |||| | | |::| | | |||| | | |:| || ||:||||| || |:|
Db   997 gcttgtaccgwgcacctggaaaacgcccagacctaaacgcccgtgaaagcagcca 1056

Qy   1489 ggnatggaarwsngtnccntayaaracngtngtcgngary.tnacnaaracngtngngnath 1548
          || | ||| | ||| | |:| | || || | |||: || | | || | |:|||
Db   1057 ggagtggggactataccctgtgaagccacagggcgagagctgccaagactaagggaac 1116

Qy   1549 tayytnycaytgycayvgayytngaygtgnmgncaygggngtnaarmgnaaytyggn 1608
          ||::|| || | || | ||:|| |:| |::| | :|| | || | :| | ::||
Db   1117 tacctctgtcatcatgngacctgnatgtgagacatggagtcagaggagaaatcattttgna 1176

Qy   1609 gcnytnmgnttygaytgycacacngtngayatgtnyngnacngacctnaarccnytntgy 1668
          | |:||:|| ||:| || | || | | | :|:: |:| | |:| | || | ::|:| |
Db   1177 acgntataatttgactggctcgctggatttgtgagacatgggggcttgaacgcctttt 1236

Qy   1669 ttyg-gncarttyttyccnttyggnacngcngtntnytacnc 1708
          :|:| |||||| |::| || ||| || |:| | | |:| ||
Db   1237 gttgtggcagttctccccatttggaatggctgtattaccc 1277

RESULT 15
AAS87262/c
ID AAS87262 standard; cDNA; 1580 BP.
XX
AC AAS87262;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23066.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

```
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
PI
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG23075.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 23066; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful for treating
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1580 BP; 357 A; 412 C; 328 G; 483 T; 0 other;
```

```
Query Match             23.1%;  Score 400.4;  DB 23;  Length 1580;
Best Local Similarity   42.4%;  Pred. No. 2.6e-92;
Matches  466;  Conservative  215;  Mismatches  383;  Indels  34;  Gaps  4;
```

```
Qy    661 aaytaytaywsnwsnttygayytngaryycaytgyaaygtnaaythgytnatg 720
Db   1107 AATTCCGATAAACGAACAAGACTGTGGCATGCTAACTAGTTACGTACGGGCGCAGTCA 1048

Qy    721 mgnaarcarathytngcnaayaargararathwsnaarcacarwsnathcargargtn 780
Db   1047 GCATCCCCCAACTTCTTAGGGACAAGTGGCATTCAGCCACCTGAGATTGAGCAATAAC 988

Qy    781 acntgggtnytnaaygcnttywsnttyathngnargcngarcayaarwsnwsngar 840
Db    987 AGAGTGTTGGTTTGGAATTGAAACTTATATTTATTTTAAAAGGATGAAAAATTGGA 928

Qy    841 aaytncaycycngayaaygtnathaararaaraaayccnttywsngargngnaarttyaar 900
Db    927 AATCTGCACGACATGACAATGAAGTCGCAATAGAAAAAGAAAAACCCATTTTCTCGGGGAGAAATTCAAG 868

Qy    901 ytngcngcngarathtgyathtgyaaygaryrtnaaayccncargayaaaygn 960
Db    867 CTGGTTGCAGAAAATTTGCAATAAAATAACAAGGAGCCAAATTTAATCACCAAGACAAACGGG 808

Qy    961 garaayathwsntggacntgycaramgnwsnwsncarcarwsnathaarwsnytngcntgg 1020
Db    807 AAAAAATGTTTTCAGGGACGCACGTCAGAGACCCTTATGCCACGCCTTATGCGGGCCCAG 748

Qy   1021 mgnccnmgnmgnaa------rtggttytgyggnacnggncnggnwsnytntgytygyytncar 1077
Db    747 AGGCCTAGGAGGAAAAATATGGTTTGTGGCTCGGCCCAGGACTTGCTTGTGTGCAG 688

Qy   1078 ccnmgngayytngtnccntcgytgyccngtnaaywsng----------------------- 1114

Db    687 TCTCAGGACTTGGTGCCTGCATCCCATCCTGTGTGGCTAAAAGGGCGGCCAATATACAGCTGAA 628

Qy   1115 -cngtngcnwsngarggngncnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
Db    627 ACCAATTGCTTCAGAGGGTGTAAGGCCCCAAAGCAATTTGGAGGCTTPACATGTGGTGTTGAACT 568

Qy   1174 gtnggngcnaaraarwsnmgnathgarrgtntgggarccnccnathmgnttycaraarath 1233
Db    567 GTGGGTACACAGAAGTCAAGAAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGGAGGATG 508

Qy   1234 tayggngaayccntggatgccnmgncaraaarttygcngtngngtnggnwsnwsntggmgn 1293
Db    507 TATGGAAACACCTGGAATGCCGCAGAGGTGTGGCTGCAGGGGTAGAGCCCTCATGGAGA 448

Qy   1294 acnwsnsngcnmgngtngtncaraarggnaaygtngngtnggaarccnccncaymgngtnccn 1353
Db    447 ACCTCTGCTTAGGGCAATGCAGAAGGGAAATGTGGGGTTGGAGCCTCCACACAGAGTCCC 388

Qy   1354 wsnggngcncanwsnwsnmgngcngtnamgnwsnccnccnwsnwsnmgnytncaraar 1413
Db    387 ACTGG------GACTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTTCCAGACCCCAGAAT 334

Qy   1414 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnaacngayacncaartgycar 1473
Db    333 GATAGACCAACTGACAGCCTTGGAATCGTGTACCTGGAAAAAGCCACAAAACACTCAAATGCCAT 274

Qy   1474 ccngtnaaargcngcngnnatggaarwsngtnccntayaaracacngtngtncgngarytnacn 1533
Db    273 CCTGTGAAAGCAGCCAGGAGGAGCTGTGTACCCTGCAAAAGCCACAGGGCAGAGCTTCCC 214

Qy   1534 aaracngtnggnathtayytnytncaytgyaygayytntgnayytnmgncayggngtnaar 1593
Db    213 AAGGCTGTAGAAAAACCCACCCTCTTACATCACGGTGTCCTGGATGTGAGACCTGGAGTCAAA 154

Qy   1594 mgngaycaaytyggngcnytnmgntnttygaytgyccnacnggnttymgnacntaatatgggn 1653
Db    153 GGAGATCATTTTGGAGTTTAAGATTTGGCTGCCCCACTGGATTGCATGGGG 94

Qy   1654 ccngt-nccnytntgyttyggncaarttyttyccnttyggnacngcngtntttyacncaartg 1712
Db     93 CCTGTAGCCCCTTTGTTTTTGGCCAAATTTCTCCCATCTGGAATGAGTGTGTTTATCCAATG 34

Qy   1713 yytntayytncaytgyat 1730
Db     33 CCCGTACCCCCATTTATAT 16
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 16:20:40 ; Search time 104.62 Seconds
(without alignments)
4064.152 Million cell updates/sec

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg..........gyytntaytyncaytgyatg 1731

Scoring table: IDENTIFY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 12281675Z residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	299.2	17.3	49136	4	US-09-422-869-1	Sequence 1, Appl
2	177.6	10.3	14855	2	US-08-687-080-59	Sequence 59, Appl
c 3	164.6	9.5	620	4	US-08-991-789A-29	Sequence 29, Appl
c 4	164.6	9.5	620	4	US-09-062-451-29	Sequence 29, Appl
5	140.6	8.1	152331	3	US-09-128-155-16	Sequence 16, Appl
6	135.8	7.8	14855	2	US-08-687-080-59	Sequence 59, Appl
c 7	67	3.9	721B	2	US-08-232-463-14	Sequence 14, Appl
c 8	59	3.4	2067	4	US-08-106-194-11	Sequence 11, Appl
9	51.8	3.0	1313	2	US-08-463-911-6	Sequence 6, Appl
10	51.8	3.0	4517	4	US-09-140-804-9	Sequence 9, Appl
c 11	50.6	2.9	1155	3	US-09-053-866-3	Sequence 3, Appl
c 12	48.8	2.8	2082	4	US-08-991-789A-2	Sequence 2, Appl
13	48.8	2.8	2088	4	US-09-440-325A-2	Sequence 2, Appl
14	48.8	2.8	289	4	US-09-351-414-3	Sequence 3, Appl
15	48.6	2.8	289	4	US-09-007-005-17	Sequence 17, Appl
16	48.6	2.8	1107	3	US-09-244-796-17	Sequence 17, App
17	48.6	2.8	1276	2	US-09-188-930-217	Sequence 217, Appl
18	48.2	2.8	2277	1	US-08-463-911-1	Sequence 1, Appl
19	48.2	2.8	2277	1	US-08-676-974-2	Sequence 2, Appl
20	48.2	2.8	2277	1	US-09-098-487-2	Sequence 2, Appl
21	48	2.8	2277	2	US-08-676-974-2	Sequence 2, Appli
22	48	2.8	2277	1	US-09-098-487-2	Sequence 2, Appli
23	48	2.8	2277	1	US-08-676-974-2	Sequence 2, Appli
c 24	44.8	2.6	7218	2	US-09-092-770-18	Sequence 18, Appl
c 25	44.4	2.6	1212	2	US-09-222-851-18	Sequence 18, Appl
c 26	44.4	2.6	1212	4	US-08-257-073-10	Sequence 10, Appl
c 27	43.4	2.5	5181	1	US-08-257-073-10	Sequence 10, Appl
c 28	43.2	2.5	729	4	US-09-140-804-10	Sequence 10, Appl
c 29	43	2.5	1215	2	US-09-092-770-8	Sequence 8, Appli
c 30	43	2.5	1215	4	US-09-222-851-8	Sequence 8, Appli
c 31	42.4	2.5	1015	4	US-09-188-930-30	Sequence 30, Appl
c 32	42	2.4	2067	4	US-09-106-194-11	Sequence 11, Appl
c 33	42	2.4	35060	4	US-08-814-095-7	Sequence 7, Appl
c 34	41.8	2.4	1001	3	US-09-188-930-218	Sequence 218, App
35	41.8	2.4	2265	3	US-09-369-618-3	Sequence 3, Appl
36	41.8	2.4	2265	3	US-09-369-617-3	Sequence 3, Appl
37	41.6	2.4	510	2	US-08-934-959-7	Sequence 7, Appl
c 38	41	2.4	597	4	US-09-528-760A-3	Sequence 3, Appl
c 39	40.8	2.4	723	3	US-08-911-423-5	Sequence 5, Appl
c 40	40.8	2.4	1659	1	US-08-231-729B-2	Sequence 2, Appl
41	40.6	2.4	2949	4	US-09-412-554A-3	Sequence 3, Appl
42	40.2	2.3	500	3	US-09-141-000-2	Sequence 2, Appl
43	40.2	2.3	1659	1	US-08-231-729B-1	Sequence 1, Appl
c 44	40	2.3	2088	3	US-09-351-414-3	Sequence 3, Appl
c 45	39.8	2.3	2265	3	US-09-369-618-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-422-869-1
Sequence 1, Application US/09422869
Patent No. 6235481
GENERAL INFORMATION:
 APPLICANT: POLONSKY, KENNETH S.
 APPLICANT: HORIKAWA, YUKIO
 APPLICANT: ODA, NAOHISA
 APPLICANT: COX, NANCY J.
 APPLICANT: SREENAN, SEAMUS
 APPLICANT: ZHOU, YUN-PING
 APPLICANT: OTANI, KENICHI
 APPLICANT: HANIS, CRAIG L.
 APPLICANT: BELL, GRAEME I.
 TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
 FILE REFERENCE: ARCD:307
 CURRENT APPLICATION NUMBER: US/09/422,869
 CURRENT FILING DATE: 1999-10-21
 EARLIER APPLICATION NUMBER: 60/134,175
 EARLIER FILING DATE: 1999-05-13
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 1
 LENGTH: 49136
 TYPE: DNA
 ORGANISM: Human
US-09-422-869-1

Query Match   17.3%;   Score 299.2; DB 4; Length 49136;
Best Local Similarity   41.1%; Pred. No. 9.8e-74;
Matches   388; Conservative   186; Mismatches 307; Indels 62; Gaps 5;

```
Qy    798 rgcnttywsnttyathmgnargcngarcayaarwsnwsngarayayytncayccngayaa 857
          :|:  |||:||| ::| | :||  ||:||::|| ||| :||:::||::|| |:| 
Db  47384 ggaactcacatttaaaaggyaagcaggagcacataaaagttaagaaattgcatcctgacta 47443

Qy    858 ygtnathaaraaraaraayccnttywsngargghaarttyaarytngcngnngarathtg 917
          :|: :| |:|| |:||||| :| ||:|| ||| || |:||| ||::||:||:|| :||:|
Db  47444 tgggatagaaagaaaaaactcattttctgaggagagaatctcaagctgcagaaaattg 47503

Qy    918 yathtgyaygaygargaryttnaaygtnaayccnacargayaydyggngaraayathwsntggac 977
          :||:  ::||  :|:| :|||:|: :||  |||  | || | || :||:: | :||  | ||
Db  47504 cataagtaactaggaagccacacatgttaatacatagacaatgaggaaaatgtctccaggggc 47563

Qy    978 ntgycarmgnwswsncarcarwsnathaarwsnytngcntggmgnccnmgnmg--naa 1034
          :||:||||| :|| :||:| | :||| |  ||  | | |||:| |:| | | :|  |
Db  47564 atgtcagaggtcttcacagacaacccaccaccatcacaggcctggaggctaggagggaaaa 47623

Qy   1035 rtggttytgyggnacnggncnggncmgnwsnsnytntgytcgygtncarccnmgnaytcc 1094
          |:|| ||:|| :|| :|| :||:||  | | :|| | :  |:||  | ::| | :| ||::
Db  47504 atgtcagaggtcttcacagacaacccaccaccatcacaggcctggaggctaggagggaaaa 47623
```

```
Db  47624  atggtttctgtcgttggggcccaggccttgctggttgtgcagtctcaggacttggtgcc  47683
           :||||||||  ||   ||    ||  |||    ||||  ||  ||| ||| || ||
Qy  1095  ntgyytnccngtnaaywsngcngtngcnwsngarggnggngcnwsnccnaarccntggcaryt  1154
           | ::| ||  ||:|
Db  47684  ccacatcc------------cagcagtggctaaaaggggccaatgtacagcttagacctt  47731
                    |:| |
Qy  1155  nccnwsnggngtngarccngtnggngcnaaaarwsnmgnathgargtntgggarccncc  1214
           |::: ||         ||:||| |  |:|:|
Db  47732  tgcttcagagtgcaagc----------------------------------------cc  47751
           ||
Qy  1215  nathmgnttycaraarathtayggnaayccntggacyancaaarttygcngtngtngg  1274
           || : |||  ::  ||  | | ||  |||
Db  47752  caagccttggtgcttacatggtgttgggcctgcagatacacagaagtttgctgcact  47811
           |::||  ||        ||  ||  :||| |:|  || | ||::|
Qy  1275  ngtngghnwsnwsntggmgnacnwsngcnmgnttgtncaraargnaaytgnggntgga  1334
           | ::||    ||  || | | :|| ||  ::|| || || | :|||||  |||
Db  47812  ggtggaaccctcaatgtagaacctctgtaggcagtgtagaagtgatatgtgggttgga  47871
            |::|| :|    |||        |     |||  ||  ||| :|  |||
Qy  1335  rcncnccncaymgngtncnwsnggnngcnc--------cnwsnsnmgngcngtnmgnmgnws  1388
           ||::|| || :|   ||  | |                  | || | ||  |||||| |
Db  47872  gcccccccacacaatcccacnggggcactgcctactgctactggaactgtgagaagaag  47931
           || | |  | || ||  |||     |  || || :|  | || || || || ||::
Qy  1389  ncchcnccnwsnsnmgnghytncaraargngmgnwsnacngaywsnytncarcaymgtncnga  1448
           ||    || || | |||   || || ||||   || || :| ||:|| | || || |
Db  47932  gccacatcctccagaccnccagaatggtagatccactgattgccttgaccactgcacctg  47991
           ||||||||||  || ||||   |||:| ||  |||| | |||  |||| ||  |||
Qy  1449  raarwsnacngayacncaartgycaccnggnatggarwsngtnccnta  1508
           ||| ||| |  ||  |  |||| ||  | || ::|||| |||:||
Db  47992  aaaagccacaagacactcaaacacacgctgtgaaggcagctgggccccttatggaaaaacctctgcta  48051
           ||:||  |||  | |  |:|  :| |  || |||  ||  | ::|| | |
Qy  1509  yaaracngtngtncngaryntnacnacacntggnnathtayytnnytcaytgycayga  1568
            :|::|||        | ||| | | |  ||    |  || || |  ::|| ||
Db  48052  caaaacaacagaggcagagctgcccaaggtcatgggagcccacactcttgcatgagcctga  48111
             ||:|| || ||  | |  |  |||:|  | | | |: ||  ||  | | || ||
Qy  1569  yytngaygtnhmgncayggngtnaaarmgngaycaytyggngcnytnmgnttygaytgycc  1628
           |||: |||  || ||  ||  ||||:||  || | |::||   :||| |||| ||:|
Db  48112  cttgaatgtgagacatggagtcaaagagatcattttggagcttaagttttgactgccc  48171
           ||:||  |||   |:||  ||| ||||| | || |:|  |  ||||| ||:|||
Qy  1629  nacngnttymgnacntayatgggnccngt-nccnytntgyttygngncarttyttyycent  1687
            || | |||    |  |:| || ||  |  ||  ||| |||  | |:   || || ||:|
Db  48172  acctgattycggacttgcatgggcctgtggccccttcattttggccaatttatcccat  48231
            |:|| || |  |||||  |||| |:|||| :||| || |||||   ||| || |||
Qy  1688  tyggnacngcngtnttyacncaartgyytntatayytncaytgyat  1730
           |::|  |  ||     || |  |   |: ||  |  || |||
Db  48232  ttggaatgggtatattacccaatgyctgtaccccattctat  48274
```

Query Match   10.3%; Score 177.6; DB 2; Length 14855;
Best Local Similarity   44.4%; Pred. No. 2.2e-39;
Matches   221; Conservative   90; Mismatches 172; Indels   15; Gaps   4;

```
Qy  1244  cntggatgccnmgncaraaarttygcngtngngtngggnwsnwsntggmgnacnwsngcnm  1303
           |||||||||   || ||  | ||   | ||   :|||  |||||
Db  14854  CCTGGATGCCCAGGCAGAGATGTTTGCTGCACGGTGTGGGGCCCCTTATGGAAAAACCTCTGCTA  14795
Qy  1304  gngtngtncaraarggnaaygtnggntggggarccncncncaymgngtncncnwsngngcnc  1363
           |||||||| ||  ||| ||    ||| ||||   |  | ||  || || |     ||
Db  14794  GGGCAATATGGAAGGGAAAATGTGGGGTTGAAACCCCAC--AGAGTTCCTATGGAGGGGAC  14737
Qy  1364  cnwsnwsnmgncngtnmgnmgnwsnytncarcaymgtncnga  1413
           ::     ||| |:|| ||||||| :|| | || ||| ||
Db  14736  TGCCTAGTGGAGCTGTGGAGAAGACAGCCACTGTGCTTCTCCAGACTGGTAGATCCACTGAAT  14677
Qy  1414  ggnmgnwsnacngaywsnytncarcaymgtncngaraarwsnacngayacncaartgycar  1473
            || |  ||| || || | |  | || ||| ||||   || | ||   ||  |  ||||
Db  14676  AATAGATCCACTGACAGCTTCCACTGTGCTTGAAAACCCCAC--AGAGTTCCTAPGGAGGGGAC  14617
Qy  1474  ccngtnaargcngcngngatggarwsngtnccntayaaracngtngtngcngarytnacn  1533
            |:|| ||  |  |||| |||   | | || |  | | || ::||| | | ||
Db  14616  CCTGTGAAAAACAACAGGAAGGAGGCTATACCCTGCAAAGCC--AGAATGGGAGCTGCCC  14559
Qy  1534  aaracngtnggnathtayytnaytncaytgycaygayytngayghtnmgncayggngtnaar  1593
           || || || |  |:| || |  | |  |||  | | | ||   ||| || || ::|||
Db  14558  AAGGCCATGGAAGCCCACCTCTTGCATCAGAGTGACCTGGATGTGAGACATGGAGTCAAA  14499
Qy  1594  mgngaycaytyggngcnytnmgnttygngncanacnggngtymgnacntayatgggn  1653
           |||  | | |::|| | :||  || | | |  |  |||:| | |  | |:| |||
Db  14498  GGAGATCATTCTGGAGCTTAAGTACACCTGCCCCACTGAATTCGGACTTGCACGGGG  14439
Qy  1654  ccngt-nccnytntgyttyggncarttyttyycenttayggnacngcngtnttyacncaartg  1712
           |:|| ||  || |||  |:||| |||| || ||:||:||  |  ||  |  ||
Db  14438  CCTGTAGCCCCTTGTTTTGGCCAATTTCTCCATTTGGAATGGGCTGTATTTGCCCAATG  14379
Qy  1713  yytntayytncaytgyat  1730
           |:|| ||  || |||
Db  14378  CCTGTATCCCCATTGTAT  14361
```

RESULT 2
US-08-687-080-59/c
; Sequence 59, Application US/0867080
; Patent No. 5965427
GENERAL INFORMATION:
  APPLICANT: Gregory Dolganov
  TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
  NUMBER OF SEQUENCES: 175
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Dehlinger & Associates
    STREET: 350 Cambridge Avenue, Suite 250
    CITY: Palo Alto
    STATE: CA
    COUNTRY: USA
    ZIP: 94306
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/687,080
    FILING DATE: 17-JUL-1996
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:

RESULT 3
US-08-991-789A-29
; Sequence 29, Application US/0899178 9A
; Patent No. 6225054
GENERAL INFORMATION:
  APPLICANT: Frudakis, Tony N.
              Smith, John M.
              Reed, Steven G.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TREATMENT AND DIAGNOSIS OF BREAST CANCER
NUMBER OF SEQUENCES: 292
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed IP Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,789A
FILING DATE: 11-Dec-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E. R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 210121.419C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 620 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-08-991-789A-29

Query Match   9.5%;   Score 164.6;   DB 4;   Length 620;
Best Local Similarity   40.7%;   Pred. No. 4.3e-37;
Matches 250; Conservative 105; Mismatches 243;   Indels 17; Gaps 6;

```
Qy   1126 gargggngcnwsncccnaarccntggccarytnccnwsnggngtngarccngtnggngcnaar 1185
          ||:||  |.| :| ||||:|| ||||| |||:|| :||  ||| || ||:|| || || ||:
Db      3 GAGGGTACCAGCCCCAAGCCTTGACAACTTCCATAGGGTGTTCAAGCCTGTGGGTGCACAG 62

Qy   1186 aarwsnmgnathgargtntgggarccnccnccnathmgnttycararaththtaygg------ 1238
          |:||| |||||| ||||:|| || ||||:|| || ||:|||||||:||||| ||||
Db     63 AAGTCAAAAATTGAGTTTTGGGATCCTCAGCCTAGATTTCAGAGGATATAAAGAAAACACC 122

Qy   1239 naayccntggatgccnmgncararartttygcngtnggngtnggnwsnsntggmgnacnws 1298
          |:|| ||| |||| || |:||:|| ||||:|| || ||||| ||||| || ||:|||||
Db    123 TAACACCTAGATATTCAGACAAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC 182

Qy   1299 ngcnmgngtngtncaraarggnaaygtnggngtnggngtnccnwsngg 1358
          || |:||:|| |:|| |||:| ||:|| |||||| ||||| ||:|| |
Db    183 TACTAGGAAAGTACAGAAGAAGAAAATGTGGGTTTGGAGCCCCAAACAGAAATCCCCTCTAG 242

Qy   1359 ngcnccnwswsnmgngcngtnmgngmngnwsncccnwsnwsnmgnytncararaarggnmg 1418
          || :| |.| ||.|:|| |:||| || | |:| ||| |:|| || || ||| ||:| |:|
Db    243 AACACTGCCTAAATGAAACTGTGAGAGAATGGCCACTGTTCATCCAGAGACACCAGAATGATAG 302

Qy   1419 nwsnacngaywsnytncaracayg tnccngaraarwsnacngayacncartgycgarcc-- 1475
          |::|| ||| | ||||:|:| ||| |:|| | |:|| || ||:||  || |:|||| ||
Db    303 ACCCACCAAAAACTTATGCCATATTGCCTATAAAACCTACAGACACTCAATGCCAGCCCC 362

Qy   1476 -ngtnaargcngcngnatggarwsngtnccntayaaracngtngcngaryatnacna 1534
          |:|||| |.|| || :| |||::| ||||:|| || |:| ||| |||||:| |||
Db    363 ATGAAAAAAAACTGAGAGAAGAGACTGTNCCCTACAATGCCACCGGAGCAGAAACTGCCCC 422

Qy   1535 aracngtnggnathtayytnytncaytgcaygaytgngayg tnmgncayggngtnaar 1593
          |:| | || ||||| |:| |:|: ||:||:| | |:||:| :||| || ||:||
Db    423 AGGCCATGGAAGCACAGCTCTTATATCAAATGTGACCTGGATGTTGAGACATGGAATCCNA 482

Qy   1594 mgngaycayttyggngncnytnmgntt---ygaytgyccnacngngnttygngnacntayat 1649
          ||:|| ||:|| ||| | |:| |||:|    |:|:| |||||| |:|::| ||| ||:| |
Db   1594 ...
```

```
Db    183 TACTAGGAAAGTACAGAAGAGAGAAATGTGGGTTTGGAGCCCCCAAACAGAATCCCCTAG 242

Qy   1359 ngcnccnwswsnomgngngcngtnmgngmgnwsncnccnwswsnmgnytncaaraarggmg 1418
          | ||  | || |  ||  |  || ||| | ||| | | |||||  || | |||||:|
Db    243 AACACTGCCTAATGAAACTGTGAGAAGATGGCCACTGTCATCCGACACCAGAATGATAG 302

Qy   1419 nwsnacngaywswsnytncarcayg-tnccngaaraarwsnacngayacncgaartgycarcc-- 1475
          | ||: || :| || | ||| |||:|| |||:||| ||| | || |||||||||
Db    303 ACCCACACAAAAACTTATGCCATAATTGCCTATAAAAACTTACAGACACTCAAATGCCAGCCCC 362

Qy   1476 -ngtnaargcngngnatggarwsngtncntayaaracngtngtncgnaarytnacna 1534
           ||:   | || ||| || |  ||| ||| || | | ||:|| || || ||
Db    363 ATGAAAAAAAAACTGAGAAGAAGAGAAGACTGTNCCCTACAANTGCACCGGAGCGAGAACTGCCC 422

Qy   1535 aracngtnggnathtayytnytncaytgycaygayytngayg-tnmgncayggngtnaar 1593
          | | || |  | ||| | | |  |||| | | | || |  ||:|| | ||| |
Db    423 AGGCCATGGAAGCACAGCTCTTATATATCAANTGCAGTTGACACATGGGAATCCNA 482

Qy   1594 mgngaycayttyggngngcnytnmgntt----ygaytgyccnacnggngnttymgnacntayat 1649
          |||:|:|||  || |  | | |   |     | || |:| | | |||| |:|||  ||
Db    483 NGAAATCNTTTAANACTTCCACGGTTNAATGACTGCCTATTANATTCNGAACTTANAT 542

Qy   1650 g-ggnccngtnccnynytntgyttyggncaarttytyccnttyggnacngngtnttyacnc 1708
          | ||  |  ||| |  || | |||  ||| |  | || |: |||| || | || |
Db    543 CCNGGCCGTGTGACCTNCTTTGCTTTGGCCATTCCCCTTTTTGGAATGGCTNTTTTTTCC 602

Qy   1709 artgyytntayytnc 1723
          :| ::|  ||:::|
Db    603 CATGCCTGTNCCCTC 617
```

```
RESULT 5
US-09-128-155-16
Sequence 16, Application US/09128155
Patent No. 6117654

GENERAL INFORMATION:
  APPLICANT: Pan, Yang
  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
  TITLE OF INVENTION: AND USES THEREOF
  FILE REFERENCE: 09404/052001
  CURRENT APPLICATION NUMBER: US/09/128,155
  CURRENT FILING DATE: 1998-08-03
  EARLIER APPLICATION NUMBER: US 60/091,650
  EARLIER FILING DATE: 1998-07-02
  EARLIER APPLICATION NUMBER: US 60/054,646
  EARLIER FILING DATE: 1997-08-04
  NUMBER OF SEQ ID NOS: 18
  SEQ ID NO 16
  LENGTH: 152331
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)..(152331)
  OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16
```

```
Query Match     8.1%;  Score 140.6;  DB 3;  Length 152331;
Best Local Similarity    32.8%;  Pred. No. 1e-27;
Matches  324; Conservative  168; Mismatches  426; Indels   71; Gaps    6;

Qy    771 hcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaa 830
          | ||:| | |  ||| |  | | | || |:| | |||| |  | | ||  |
Db  71225 tcagggtaaagaaatgactaaagttggaacctatgtttaaaatggaagtagagctaa 71284

Qy    831 rwsnwsngaraayytncaycncgayaaygtnathaaraaraayycnttywsngargg 890
          | :|| | | |  | |: |  | | :||  |||| ||| || | || || |
Db  71285 aaattggaaaattgcagcctggtccttgtcggagagaaagaaatccaagtaggctgcaga 71344

Qy    891 naarttyaarytngcngcngarathgyathgyayaaygaaygaarytnaaygtnaaycncca 950
          :|:|        |  | |  || || | ||||| |  |  | | |  || || |
```

```
Db  71345 gcaatcaatgctag---agagattagtcatgactaaataaagggaggccaagtgctaatattca 71400

Qy    951 rgayaaygggngaaraayathwsntggacntgycaamgnwsnwsncarcarwsnathaarws 1010
          :|:|:|| || :||||  |  |||| || | ||| |  |  || | || || |
Db  71401 agacaatgttaaaaaggcttgagggcattcagaagatctatgaagcagccccctccatc 71460

Qy   1011 nytngcntgmg------nccnmgnmgnaartgttgytyggngnacnggngwsny 1063
          | |    | |    |  ||   || |  | | || | |   || |  |||
Db  71461 acagggtgcaagaggtttggtgcactaggcccaagacgtttatgggcaanngccaggccac 71520

Qy   1064 tntgytgygtncarccnmgnaaryytngtnccntgyygtnccngtnaa----------- 1109
          |  |:|||  ||| |  | | || || | | | || || || |
Db  71521 actgctatgcacagctttgggacactgtgcccgatccagtcctctggctcc 71580

Qy   1110 --------------------ywsngcngtngcwsngargggncwsncc 1139
                              | || | || || | ||| |:| ||
Db  71581 accctgcttcaaacggccaagatagagcttggacactgctccgagggcacaagcca 71640

Qy   1140 naarccntggcarytnccnwsnggngtngarccngncnaataarwsnmgnath-- 1197
          ||| | ||| |  | ||| | |||| |: | |  |  | | | || |||
Db  71641 taagcctggttggttccactgtggttaagcctgcaggtgccaagattgag 71700

Qy   1198 gargtngtggarccncncnathmgnttycaaraarathtayggnaaayccntggatgcnmgn 1257
          || |:||| | | |  | || | |:||| | | |  || ||||| | |||| | ||
Db  71701 ggagcttgggcacttccacctcaaatttcagaggatgtgtcaggaaacctaggttcccag 71760

Qy   1258 caraarttygcngtnggngtnggnwsnwsngggmgnacnwsngcnmgntngtngncaaraar 1317
          || |  |: | | ||| | | || |  || | |:| ||| |   | || |
Db  71761 cagaagacatgatacaggggcagagaccctgcagaaaccctactaggtcaatgccaaag 71820

Qy   1318 ggnaaaytgnggntggarcncncncaymgntgtnccnwsnggngncnccncnwsnsnwsnmgncn 1377
          ||   |:|| | || |:| |    | |  || |    | | |  || | || ||| |:|
Db  71821 gaaaatgtgggtttggagtcctccacactggtccccactggtgtatact 71880

Qy   1378 gtnmgnmgnwsncncncnwsnwsnmgnytncca 1437
          || | || ||    |  | | |||  ||
Db  71881 gtgggaatgggctgctgccctccagaatggttgacca 71940

Qy   1438 caygtnccngaaraarwsnacngayacncaartgycarccngtnaargcngcnggnatggar 1497
          || ||: ||  | || |||| |  || |  ||| || |: | || |  | |:||| ||
Db  71941 cctgagcctggaaaagctgcaggcccacataggtagatgcactggcgagctggcac 72000

Qy   1498 wsngtnccntayaaracngtngcngarytnacnacnaaracngtnggngnatthtayytnytn 1557
          |:|| || || |  | ||| | |  | |   | || | | | || || |  || | |
Db  72001 ac--tcccagaggaagccacacagaggcaggccttggggagcctacccctg 72058

Qy   1558 caytgycaygayytngayghtngmgncaytgyggngncytnmgn 1617
          || | | ||  | ||  |  |:| |  || |:  |  | |  ||
Db  72059 aaccagcttgcaggacgtggaatcaa-----agatatgtgcagcttaag 72106

Qy   1618 ttygaytgyccnacngnnttymgnacntayatgggncgnccnytnccnytgtyttyggncar 1677
          || |:|  || | | | | ||  || ||: |||| |  | | |||:| || | ||
Db  72107 cttaatgtttccctgtcaattcacgcttgtgtggaccctgctttttttt 72166

Qy   1678 ttyttyccnttyggnacngcngtnttyac 1706
          || ||  | || |   || | | |||
Db  72167 tttttttttttggtcacaggtgtttgaac 72195
```

```
RESULT 6
US-08-687-080-59
Sequence 59, Application US/08687080
Patent No. 5965427

GENERAL INFORMATION:
  APPLICANT: Gregory Dolganov
  TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
  NUMBER OF SEQUENCES: 175
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Dehlinger & Associates
    STREET: 350 Cambridge Avenue, Suite 250
    CITY: Palo Alto
    STATE: CA
    COUNTRY: USA
```

```
;  ZIP:  94306
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE:  Floppy disk
;    COMPUTER:  IBM PC compatible
;    OPERATING SYSTEM:  PC-DOS/MS-DOS
;    SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/08/687,080
;    FILING DATE:  17-JUL-1996
;    CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US 08/592,126
;    FILING DATE:  26-JAN-1996
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  Sholtz, Charles K.
;    REGISTRATION NUMBER:  38,615
;    REFERENCE/DOCKET NUMBER:  4600-0111.30
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  (415) 324-0880
;    TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH:  14855 base pairs
;    TYPE:  nucleic acid
;    STRANDEDNESS:  double
;    TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;    INDIVIDUAL ISOLATE:  5' END OF INTRON 2 OF RAD50 GENOMIC
;    INDIVIDUAL ISOLATE:  SEQUENCE
US-08-687-080-59

Query Match            7.8%;  Score 135.8;  DB 2;  Length 14855;
Best Local Similarity  39.5%;  Pred. No. 2.2e-27;
Matches 191;  Conservative 101;  Mismatches 160;  Indels 31;  Gaps 3;

Qy  798 rgcnttywsnttyathmgngargncngaracayaarwswsngaraayytncayccngayaa 857
         |:|||:|:  ||||:  |  |:|| ||:|:||||:|||||:  |::|  || |:|:|:
Db 5187 GGAACTTATATATTTAAAAGGGAAGCAGAGACATAAAAAGTTTGGAAAAATTTGCACCCTGATCA 5246

Qy  858 ygtnathaaraaraaraaycnttywsngargngnaarttyaarytngcngcngarathtg 917
         ::||:|  |:|  |||  ||||  |:|||||||:||||||::||:  ||::|:|||||||
Db 5247 TGTGGTAGAAAAAGAAAAACCCATTTTCTGGGGAGGAAATTCAAGCCGGCGTCCAGAAATTTG 5306

Qy  918 yathtgyaaygargngarytnaayg-tnaayccncargayaaygngaraaythwsntgga 976
         :|:||:|||:|||||||::||:|||  ||||||:||||||||:|||||||||:||:||||
Db 5307 CATAAGTAAAGAAGGAGGACAGAAGTGTTAATAGCCAAGACACTGGGGAAAATGGCTAGAAGG 5366

Qy  977 cntgycarngnwswsncarcarvsnathaarwsnytngcntgngncnmgmgnnaart 1036
         |:|||:|||:|||||:  || |::|  |::| |||||::|||:|:||  |:| ||||
Db 5367 CATTTCAGAGATCTTCACACAGCCCCTCCCATTACAGGACACTACTCCC--------- 5416

Qy 1037 ggttytgyggnacngngnvwsnytntgyytgyytncarccmngngayytngtnccnt 1096
         ||  ||:| :|||  |:  ||  | ||:  |:: |  |:|  |::|  ||  ||::|
Db 5417 TGTGTTTCAGCCACTCCACCATGGCTAAAAGGCCCCAGATATATC----- 5469

Qy 1097 gygtnccngtnaaysngcngtngcnwsnccnaarccntggcarytnc 1156
         ||:||||| ::|  || |||  |||  |||  ||||||:|||:||
Db 5470 --------TTAAGCTGCTGCTCCACAGGCGTGCAGCCTTGGCAGCTTC 5516

Qy 1157 cnwsnggngtngarccngtnggngncnaataraaarwsnmgnathgargtntgggarccnccna 1216
         ||  ||||:||||||||::||||||||||||:  || |||||||:|||:||  |||||  ||||
Db 5517 CACATGGTGTTAAGCCTGGGGTGCCGCAGAGAACAAGAGTTGAGGCTTGGGAGCCTCCAC 5576

Qy 1217 thmgnttycaraararathtayggnaayccntggatgcnnmgncaraarttygcngtnggng 1276
         :||  |:|||:|  |||:|||  |||||||||:|:  |||:|||||||  |:||:|||:||
Db 5577 CTAGATTTCAGGGGATGATATGGGAAAACGTCTGGAGCAGAGAAGTCTGGAGACTTG 5636

Qy 1277 tng 1279
         |:|
Db 5637 TTG 5639
```

```
RESULT  7
US-08-232-463-14/c
;  Sequence 14, Application US/08232463
;  Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE:  Foley & Lardner
;    STREET:  1800 Diagonal Road, Suite 500
;    CITY:  Alexandria
;    STATE:  VA
;    COUNTRY:  USA
;    ZIP:  22313-0299
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE:  Floppy disk
;    COMPUTER:  IBM PC compatible
;    OPERATING SYSTEM:  PC-DOS/MS-DOS
;    SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER:  US/08/232,463
;    FILING DATE:
;    CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  US/07/935,313
;    FILING DATE:  26-AUG-1991
;    APPLICATION NUMBER:  EP 91 114 300.6
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  BENY, Stephen A.
;    REGISTRATION NUMBER:  29,768
;    REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  (703)836-9300
;    TELEFAX:  (703)683-4109
;    TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH:  7218 base pairs
;    TYPE:  nucleic acid
;    STRANDEDNESS:  single
;    TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;    CLONE:  pT2gpt-F1s
US-08-232-463-14

Query Match            3.9%;  Score 67;  DB 1;  Length 7218;
Best Local Similarity  12.6%;  Pred. No. 5.7e-08;
Matches 64;  Conservative 203;  Mismatches 241;  Indels 0;  Gaps 0;

Qy  580 athathttyacangngtnytntayaaygcncarmgngayytnaaargncnatggngtn 639
         |::  ::|  |  |:||:|  ||:||  |:||:|| |:|:||  |  ||:|||||||:|
Db 1536 ATCGCTTCTAGACGCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTA 1477

Qy  640 ttygcntgymgngtnccngnaaytaytaywsnwsnttygavtgtngarytncaycaytgy 699
         |:|||:|:||||::||  |  ||:||:||  |  ||||:|:| |:|||:|||||::|
Db 1476 CCTATCTATGCAAGTAGTAGTTAAAGAGTAGAGAGAAATTTGGTACRARRRRRRRRRRRRRR 1417

Qy  700 aargtnaayathtggytnatgmgnaarcarathytngcnaayaarygargarathwsnaar 759
         ||:|||| :||||:  |:|:|||||||:| :||||||:|||:|||:||||| |||  ||||
Db 1416 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1357

Qy  760 carcarwsnathcaargargtnacntgggtnytnaytnaaargcnttywsnttyathmgngar 819
         |||||  |:||  |||||||:||:|  |:|:||||  |  |::|:|:  :||:| :|||
Db 1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
```

```
Qy   820 gcngarcayaarwsnsngaraayytncayccngayaaygtnathaarraaraaraayccn 879
             ||.|||     |  ||  :  ::||  |   :| | |:.|:.||       |:::| ||
Db  1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237

Qy   880 ttywsnsgargggnaarttyaaarytngcngcngarathtgyathtgyaaygargarytnaay 939
             |:|:|  :|    | : ||:: | ||:| |:  ||::||  |  :  ::||  | ||.||
Db  1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177

Qy   940 gtnaayccncacgayaaygngaraayathwsntggacntgycaarmgnwsnwsncarcar 999
             | :::| ||| | |:|  :: | | | :||  | | ||  : : |   ||  ::|
Db  1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117

Qy  1000 wsnathaarwsnytngcntggmgnccnmgnmgnaaartggttytgyggnacnggnccnggn 1059
             ||||  || | | | | ||  || :| |  : :|| : | | :  || | || ::| |
Db  1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1057

Qy  1060 wsnytntgytgygtncarccnmgngayy 1087
             | |  | |  | | |  |  : :|
Db  1056 CCCTCGACCTGCAGCCTCGGAATT 1029
```

```
Qy   425 cnmgncccncganrcgaargtnggnccncaggngcgngnytgccncaartayacngnq 484
             |   |:| || :| : :|        :|  |        |::  |  |  || |
Db    13 YTNGARMGNCCNCCNGGNGCGARGTNWSNYTNGCNGCCNTTYTYTNACNGARCAR 72

Qy   485 arathwsngaratgacnaartgyccntgyccngayathgaymgnwsngcnttyacngtna 544
             |:|:|| | || :| | : |  |  | |:   | | | :  ||     | | :|  ||
Db    73 AARTTYACNYTNTGGATHMGNCARGCNCARATGWSNTGYGARAAYGCNCAYGCNYTNCCN 132

Qy   545 arytnwsngnaarytnccnytnccnttyaarccnaathathttyacnggngtnytntaya 604
             |  | |  | :  | |  :  |   | |     |  :| | | : : |    | |  :|
Db   133 WSNACNCNCCNTTYTNTTYTTYMGNGNRAARCNAAYCCNCCNACNCCNCNGNYTNCCN 186

Qy   605 aaygcncacamgngayytnaargarcgnatgggngtnttygcnggngtnccnggnaayt 664
             ||  | ||| | | :|   ||:|  |  |  | :|  |:   | |  | | ::| |  |
Db   187 WSNCCNWSNNSNYTNTTNACNAAARCCNCCNACNCCNCNGNYTNCCNGNCNYTNCCN 246

Qy   665 aaytaywsnwsnttygayngtngaryytncaycaytgyaaygtnaayathtgyytnatgmna 724
             ||  | |   | | :|   |  | ||   | :| |  |  | || | : | :  | || |
Db   247 CCNCNCCNCCNCARWSNNAARWSNCARAARAAYSNWSNGARCCNGARGCNTGNCNTTYTY 306

Qy   725 arcarathytngcnaayaargargarathwsnaarcarcarwsnathcargagtnacnt 784
             | || | |    :||| |:| | | |   ||  |   | ||| | | || | | |  ||
Db   307 AARAARCARGCNCAYTGGMGNGARMGNAARMGNAARACNAARCCNAARCARAAYCCNGGN 366

Qy   785 gggtnytnytnaargcnttywsnttyatthmgnargcngarcaycayaarwsnsngaraayy 844
             | | | | |:|| |  |:| || | ||.|.:| : | :| |||  | |  ||  || |:|
Db   367 ACNMGNCARCCNGARCAYTTYTYTYTYTTYTTYACNYTNCCNGARAAYGARAAYARCAR 426

Qy   845 tncaycCngayaaygtnathaarraaraaraaycCnttywsngargggnaarttyaarytng 904
             | | |  | || | ||    |   |   |  |  | | || | || : | |  ||: | |
Db   427 WSNWSNNAARCARWSNCCNCNCCNACNWSNAARYTNYTNACNCNGNMGNGNMGNCARACNWSN 486

Qy   905 cngcngarathtgyaayttgyaaygtnaaytgnaaytgnaaygtnaayccncaargayaaygngara 964
             ||| | | | | :|  | || |  | | :| |  |  | |  | | ||  |  |: | || | |
Db   487 CCNTGYWSNCAYGGNGTNMGNYTNAARGGNTGGTTYWSNMGNCARWSNWSNWSNCCN 546

Qy   965 ayathwsntggacntgycarngnwsnwsncarcarwsnathaarwsnytngcntggmgnc 1024
             |: |: ||:| | | | |   | |   ||:||   ||  |  |   | | ||.| |.|:
Db   547 ACNACNTTYTYTYACNMGNCNGNGARMGNGAYTTYGGCNYTNMGNGGNWSNGGGGAR 606

Qy  1025 cnmgnmgnaartggttytgyggnacnggncCnggnaaycnttngtngtygyggtncarccnmgng 1084
             |   : | || ||    : :|  | |   : | | | | :| |  | :| | | |  |
Db   607 CARYTNCAYGAYTAYGCNGARMGNGARWSNSGNCAYACNATGMGNGNCNCARATHGARGTN 666

Qy  1085 ayyttngtncCntgygtnaccngtnaaywsncngtngcnwsngarggngngcnwsnccnaaarc 1144
             |   | | |   | ||  | :| || |    | |:  | | | ||  | :|   | |  ||  |
Db   667 ATHCCNTGYAARATHGTGGNGAYAARWSNWSNGGNATHCAYTAYG---GNGTNATHACN 723

Qy  1145 cnttggcarytnccnwsnggngtngarccngtnggggcnaaaaaarwsmgnathgargtnt 1204
             ||  | | || | ||  || | :|   : || |||  |:| | | |  |: | |: |
Db   724 TCYGARGGNTGYAARGGNTTYTTYMGNMGNWSNCARCARAAAYAAYGCNWSNTAYWSNTGY 783

Qy  1205 ggggarccncCnathmgnttycaraaraathtayggnaaayccntggatgccnmgncaraart 1264
             || | | ||  | | || |  |     | |  |:| | | | |   || :| |  || | |
Db   784 CCNMGNCARMGNAAYTGYYTNATHGAYGGNACNAAAYMGNAAYMGNTGYCARCAYGTYMGN 843

Qy  1265 tygcngtnggngtnaggnwsnwsntggngnaacnwsngcnmgngtngtncaaraaggnaayg 1324
             | ||| |:| |:| |  | |   | ::| |  |  || :| |  : |  |:|| | |  |
Db   844 YTNCARAARTGYTGNGCNYTNGCNGTNAAGGNAYGCNGTNAARTTYGGNMG---NATG 900

Qy  1325 tnggntgggarccncncacamgngcnccnwsnsmngngcngtnmgnm 1384
             | ||:| ::| | |   ||| | |   || || |  :| ||  |||
Db   901 WSNAARARACARCNGNGAVWSNYNTNTAYGCNGARGTNCARARARCAYCARCARMGNYTNCAR 960

Qy  1385 gnwsncccnccnwsnmgnytncaraaggnmgnnacngaywsnytncarcayngtnc 1444
             | | || |   ||  | | |    :|  :  | | || ||   | | |  | :|
Db   961 GARCARMGNCARCARCARWSNGGNGARGCCNGARGCNYTNGCNGTNTAYWSNWSNWSN 1020

Qy  1445 cngaraarwsnacngaycncartgyaarccngtnaargcngcngngnatggarwsngtnc 1504
             ||:|  |.|  :| | || |.|   | || :| |  | |  |  | || | | | || |
Db  1021 ATHWSNAAYGGNTTNWSNAAYTNAAYAYNAAYGARACNWSNGGNACNTAYGCNAAYGGNCAY 1080

Qy  1505 cntaya 1510
```

Db 1081 GTNATH 1086

RESULT 9
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 73..804
US-08-463-911-6

Query Match   3.0%; Score 51.8; DB 2; Length 1313;
Best Local Similarity 34.2%; Pred. No. 0.00022;
Matches 75; Conservative 36; Mismatches 108; Indels 0; Gaps 0;

```
Qy  512 gyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnccnytnccnt 571
Db  392 GTGCCTARGTAYACCGCTCAGCATTCACTGTGGGATTCGGAGACTTACGTTACTATCCCCA 451

Qy  572 tyaarccnathathttyacnggngtnytntayaaygcncarmngngayytnaargargcna 631
Db  452 ACATGCCCATTCGCTTTACCAAGATCTTCTACAATCAGCAAAAACCACTATGATGGCTCCA 511

Qy  632 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytnc 691
Db  512 CTGGTAAATTCCACTGAACATTCCTGGGCTGTACTACTTTGCCTACCACATCACAGTCT 571

Qy  692 aycaytgyaargtnaayathtggytnatgmgnaarcara 730
Db  572 ATATGAAGGAGATGTGAAGGTCAGCCTTCTTCAAGAAGGACA 610
```

RESULT 10
US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 4517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-140-804-9

Query Match   3.0%; Score 51.8; DB 4; Length 4517;
Best Local Similarity 34.2%; Pred. No. 0.0081;
Matches 75; Conservative 36; Mismatches 108; Indels 0; Gaps 0;

```
Qy  512 gyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnccnytnccnt 571
Db  346 gtgcctatgtataccgctcagcattcagtgtgggattcggagacttacgttactatccca 405

Qy  572 tyaarccnathathttyacnggngtnytntayaaygcncarmngngayytnaargargcna 631
Db  406 acatgcccattcgctttaccaagatcttctacaatcagcaaaaaccactatgatggctcca 465

Qy  632 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytnc 691
Db  466 ctggtaaattccactgaacattcctgggctgtactactttgcctaccacatcacagtc 525

Qy  692 aycaytgyaargtnaayathtggytnatgmgnaarcara 730
Db  526 atatgaaggagatgtgaaggtcagcctccttcaagaaggaca 564
```

RESULT 11
US-09-053-866-3/c
Sequence 3, Application US/09053866
; Patent No. 6111075
; GENERAL INFORMATION:
; APPLICANT: Xu, Wenfeng
; APPLICANT: Presnell, Scott R.
; APPLICANT: Yee, David P.
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
; TITLE OF INVENTION: PAR4 (ZCHEMR2)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053,866
; FILING DATE:
; PRIOR APPLICATION DATA:
; CLASSIFICATION:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```
NAME: Leith, Debra K
REGISTRATION NUMBER: 32,619
REFERENCE/DOCKET NUMBER: 98-10
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 206-442-6674
  TELEFAX:   206-442-6678
  TELEX:
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1155 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
US-09-053-866-3
```

```
Query Match            2.9%;  Score 50.6;  DB 3;  Length 1155;
Best Local Similarity  6.3%;  Pred. No. 0.00043;
Matches   64;  Conservative  256;  Mismatches 703;  Indels 0;  Gaps 0;
```

```
Qy    656 cnggnaaytaytawsnwsnttygaaytgyacaytgyaaygtnaaayathtggy   715
          |:|:|                                    |::|||:||||:
Db   1071 YTGRAANARNCCNGCNCNACYTTRTCNCKRAAYTCNGCNSWNACRTARTARTADATRA  1012

Qy    716 tnatgmgnaaacarathytngcnaayaargargagarathwsnaarcarcarwsnathcarg   775
          :|:|                                          | ||   :::
Db   1011 NGGRTCNACRCANSWRTTNARNGTNSWNARNGGGNACRTANGCNCNRCTANAR   952

Qy    776 argtnacntgggtnaytnytnaargcnttywsnttyathmgngargcngarcayaarwsnw   835
          ||:|                                   :|  |||
Db    951 RTTNCCCANGCNSWNGGNWGGRTCNSWRTARTGNARNARNARNARRTTNSWNGG   892

Qy    836 snqaraaaytncaycaycnqaayaaygtnathaaraaraaayccnttysnqarggnaart   895
          :|:|||                                 ||
Db    891 NACRAARAANGCNACNGCNSWNGGCNARNACNACNGCNTNARNCNNARNGGCRTGNCCRTA   832

Qy    896 tyaarytngcngcngarathtgyaathtgyaaaygyaaygtnaaayccncaargaya   955
          :|:                                      |:||:|
Db    831 NCKNCKNCNCGNCGNARNGTRTGNARNGTNGCNCCRTARCANARNARNACATNGCNAR   772

Qy    956 ayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytng   1015
          :||:|                            |::::
Db    771 NARNGGNARRAANCANCNARNARNGNARCANGTRAANGCNGGYTGCCARTGNSWNGC   712

Qy   1016 cntggngnccnmgmgnnaartggttyttygggnacnggcnggnwsnytntgytcgygtnc   1075
          |:||:|                              ||:|:|
Db    711 YTGNGCRTCNARNGGNARNGCRTCTRGRCANARNACNCKRTCNSWNCKNGCNARNCKRAA   652

Qy   1076 arccnmgngayytngtnccntgygtnccngtnaaaywsngcngtngcnwsngarggngncnw   1135
          |::                                      | |
Db    651 NGTYTGNCKYTGNARNGTNARNGCNARNGCGNGCCATNARCCANGCNGCCAT   592

Qy   1136 snccnaarccntgggcarytnccnwsnggngtngarccngtngngcnaaaraarwsnmgna   1195
          ::|:                                  |:|:|
Db    591 RCANARNCCNARNGCNCKNCCNCKNARNGCNARNGGRTGNACNAR   532

Qy   1196 thgargtntgggarccncncnaathmgntycaaraararathtaygngnaaycntggatgcnm   1255
          ::|:|                                 || |||
Db    531 NGCNARRTANCKRTCNARNRNWNACNGCNGCNARNARNARNACNSWNCCRTACATRTGNCC   472

Qy   1256 gncaraarttygcngtngngtnggngnacnwsnggncngnacnwsngcncrgtngttcara   1315
          :::|                                 ||:
Db    471 RTANARGNGCNGTNGCNARNCKCAGNGCCYTCNCCRAANGCCCANCKYTGNCCNCCK   412

Qy   1316 arggnaaygtngngntgggarccncncnacayngngtnccnccnwsnwsnmgng   1375
          :|:|                              | |::|
Db    411 NARRGRTANGCDATNCKNGGNGGNGGNGGNCNARNARNARNRRTCNGCNGTNGCNAR   352

Qy   1376 cngtnmgngnwsncncnccnwsnwsnmgnytncaraaggnmgnmgwsnacngaywsnytnc   1435
          |:|:                                 |::
Db    351 RTTCATNARNARCATNGTNSWNGGNARNACCNANARNGC   292

Qy   1436 arcaaytncngaraarwsnacngaacnacncaartgycarccngtnaaargcngcnggnatgg   1495
          |:|                                       ||:
Db    291 NARNCCRTTNGCNGGNARNCCNACNACNACNARNACNARNCCRTANARNGCNGGNACNARNCK   232

Qy   1496 arwsngtnccntayaaracngtngtncgngarytgacnaaracngtnggnaathtayytny   1555
          |:||                                 ||:
Db    231 NGTNGGNACCCANCNCANARNARNARNGNCNCKNSWNSWRTCNGGNARYTCNARNGTRTCNSW   172

Qy   1556 tncaytcycaygaytngaygtnmgncaygygghgtnaarmgnqaycaytqyqngncnytnm   1615
          |::                                  || |
Db    171 RTCRTTNGCRCANACYTGNCCNGGRTANCCNCKNGGNGCNGGNARDATNSWNGGNGTNSW   112

Qy   1616 gnttygaytgyccnacngngttymgnacntayatgggncncngtnccnytntgytgyggnc   1675
          :|:|                                 :|
Db    111 RTCRTCNCCNCCNCCNGTCNGNTGGNTCNTCNNSWYTCRTCRTANACNSWNGGNGTYTGNGTNCCNCC   52

Qy   1676 art   1678
          ::
Db     51 NSW   49
```

```
RESULT 12
US-09-440-325A-2
Sequence 2, Application US/09440325A
Patent No. 6280994
GENERAL INFORMATION:
  APPLICANT: Sheppard, Paul O.
  TITLE OF INVENTION: Zacl: A Human Metalloenzyme
  FILE REFERENCE:
  CURRENT APPLICATION NUMBER: 98-79
  CURRENT FILING DATE: 1999-11-15
  NUMBER OF SEQ ID NOS: 3
  SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2
  LENGTH: 2082
  TYPE: DNA
  ORGANISM: Artificial Sequence
  FEATURE:
    OTHER INFORMATION: Artificial Sequence
    OTHER INFORMATION: This degenerate sequence encodes the amino acid
    NAME/KEY: misc_feature
    LOCATION: (1)..(2082)
    OTHER INFORMATION: n = A,T,C or G
US-09-440-325A-2
```

```
Query Match            2.8%;  Score 48.8;  DB 4;  Length 2082;
Best Local Similarity 10.4%;  Pred. No. 0.0026;
Matches   82;  Conservative  191;  Mismatches 512;  Indels 0;  Gaps 0;
```

```
Qy    710 thtggytnatgmgnaarcarathytngcnaayaaargargarathwsnaarcarcarwsna   769
          |::||                                    | ||
Db    316 ttyggnacncargcnaygytnytttyaaargtnacncartynytyaargtnaaaygn   375

Qy    770 thcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcaya   829
          |::||                                    :|
Db    376 atgytnwsnaarytncaraaytggaaargcngncnytnwsnaargaryatnmgnar   435

Qy    830 arwsnwsngaraaytncayccngaaygtnathaaraaraaayccnttysnngarg   889
          :||:|                                 |::|
Db    436 tayaaygarytnytngcnnnnytngaratgacntaywsnatggcncnargtntgyytnaay   495

Qy    890 gnaarttyaarytgcngcngarathtgyaythtgyaaygaygargarytnaaygtnaaycnc   949
          |::|                                      |:|
Db    496 gargnccntgytynwsnytgarwsngarytnqaratqattcaacarwsnsnmgnayaar   555

Qy    950 argaraaaygggaraaayathwsntgacntgycarmgnwsnwsncarcarwsnathaarw   1009
          :|:|                                 |:|
Db    556 gargarytnytgggntggcaargtncgngtnngngncaratthgyacn   615

Qy   1010 snytngcntggmgncnmgmgnaatggttyttttyggnacnggngncaraggnwsnytntgyt   1069
          :|:                                   |::
Db    616 acnttyaarcaytayggntaggaryaawsngaayaargcgcncnaartgtnaaaygngngaar   675
```

```
Qy   1070  gygtncarccnmgngayytngtnccntgygtnccngtnagcnwsngarg 1129
           |:||  :||     |  |:   ||     ::
Db    676  gayatgggngcnytntggcayywsnaartaygarwsngayacnytngar 735

Qy   1130  gngcnwsnccnaacccntggcaryytnccnwsnggngtngarccnaaraarw 1189
           |:||:  |  ::    ::  ||  :  |  ::
Db    736  mgnytntycaargaryatnmgnccnytntayytnaayccncaytaygtnmgmgngcn 795

Qy   1190  snmgnathgargtntggarccnccnathmgntycaaraathtaygngnaayccntgga 1249
           |  |:||  |           ||:        |
Db    796  ytncaymgncaytaygnggncngarytnathgaytgnmgnggnccnathccngcncaytn 855

Qy   1250  tgccnmgncaraarttygcngtnaggngtnggngnwsntggmgnacnwsgcnmgntg 1309
           |:||     :  ||      |       |
Db    856  ytnggngaraaycnytngcncarwsntgggtnaayathytngayccngtnytnccntty 915

Qy   1310  tncararaggnaaygtngngntggnarccncccncaymgngtnccnwsnggngcncnccnwsnw 1369
           |:||:       ||  ::    |   ::
Db    916  ytnaaararathccngaragaygtnacaarathtaygngnaaycccngar 975

Qy   1370  snmgngncngtnmgnmgnwsnccncnwswsnmgnytncararaggnmgnmgnwsnacngayw 1429
           |  |:||       ||:
Db    976  aarytnaatgytngargcngaracnttyttyacntaaytytngnytngcnytnccnccn 1035

Qy   1430  snytncarcaygtnccngararaarwsnacngayacncaartgycarccngtnaaargcngcng 1489
           |:||:      |||         :::
Db   1036  gcncnccnwsnttytggaaraarytnatgytnatgmgnccnacnacangaygnggnmgnargtn 1095

Qy   1490  gnatg 1494
           ||:|
Db   1096  gargtg 1100
```

```
RESULT  13
US-09-351-414-3
Sequence 3, Application US/09351414
Patent No. 6265199
GENERAL INFORMATION:
APPLICANT: Shepperd, Paul O.
APPLICANT: Baindur, Nand
APPLICANT: Deisher, Theresa A.
APPLICANT: Bishop, Paul D.
TITLE OF INVENTION: DISINTEGRIN HOMOLOG
FILE REFERENCE: 98-29
CURRENT APPLICATION NUMBER: US/09/351,414
CURRENT FILING DATE: 1999-07-09
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2088
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: zdint1 amino acid degenerate sequence
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(2088)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(2088)
OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3
```

```
Query Match       2.8%;  Score 48.8;  DB 4;  Length 2088;
Best Local Similarity  11.8%;  Pred. No. 0.0026;
Matches 140;  Conservative 286;  Mismatches 749;  Indels 8;  Gaps 3;

Qy    544  aarytnwsngnaarytnccnytyaarccnathathttyacnggngtnytntay 603
           |  ::||       :::||   | :::||
Db    731  aratgytncaygarttywsnaartaymgncaarmgnathaarcacaycaygngcngtnc 790
```

```
Qy    604  aaygcncarmgngaytnaargcnatggggcngtntttygcntgymgngtnccngnaay 663
           |:||     :::             ||            :
Db    791  ayytnathwsmgngtnacnttycaytayaarmgnwsnsnytnwsntayytcygargng 850

Qy    664  taytaywsnwsnttygaytngaryyatnccayytgytaargtnatgmgn 723
           |:||:   ::|     |       ||    |
Db    851  tntgtywsnmgnacnmgnggngtnggngtnytncaraayytgtaargtnaaygaaygaygtnccnatggccngtngcnc 910

Qy    724  aarcarathytngcnaayargargaratthwsnaarcarcarwsnathcargargtnacn 783
           |:||      |||   :::||
Db    911  argtnytnwsncarwsnytngcncaraayytnggnnathcartgggargarccnwsnwsmgna 970

Qy    784  tgggtnytnaaargcnttywsnttyatthmgnaatcgcngacayaaarwsnwsngaraay 843
           |:||             ::|       :|
Db    971  arccnaartgygaytgyacngarwsntgggnwsntgyathatgaagaracngggtnw 1030

Qy    844  ytncayccngayaaygtnathaaraaraaaycnttywsngarggnaarttyaartn 903
           |:||     ::|            |
Db   1031  sncaywsnmgnaaarttywsnaartgytngaartaymgngayttyttyytncarmgng 1090

Qy    904  gcngcngaratgtgayathtgyaaygargaryatnaaygtnaaycnacargayaaygngar 963
           |:||            :||
Db   1091  gnggngngcntgaytgctyttyaaymgnccnacnaaarytnttygarccnacnargtyggna 1150

Qy    964  aayathwsntggacntgycarmgnsnwsnscarcarwsnathaarwsnytngcntgmgn 1023
           |:||       ::|
Db   1151  ayggntaygtnga--rgcnggngargartggyaygtyggnttycaygtnartgytayg 1208

Qy   1024  ccnmgnmgnaartggttgtygggnacnggngcncnggnwsnytntgtygytncarccnmgn 1083
           |:||          ||
Db   1209  nytntgtygyaararatgywsnytnwsnaaygggncaytggwsng-ayggncctgyt 1267

Qy   1084  gayytngtnccntgygtnccngtnaaywsngcngtngcnwsngargggncwsnccnaar 1143
           |:||
Db   1268  gyaayaayacnwsntgyytnttycarccnmggntaygartgymgngaycgngtnaayg 1327

Qy   1144  ccntggcaryytnccnwsngggngtngarccngtnggggcnaararwsnmgnathgargtn 1203
           |:||
Db   1328  artgygayathacngartaytgycacngngaywsngggncatggccccnaaayytncaya 1387

Qy   1204  tgggarccnccnaathmgnttycararaathtaygngnaayccntggatcgccnmgncaraar 1263
           |:||                       :::
Db   1388  arcaргayggntaygcntgayacangaycargggmgntgytayaaygngarttgyaara 1447

Qy   1264  ttygcngtnggngtnggmgnacnwsngcnmgngtntgtncararggnaay 1323
           |:||
Db   1448  cnmgngayaaycaartgycartayathcgggngacnaargcngcngngpwsnsnagayaarttyt 1507

Qy   1324  gtngntgggarccncacaymgntaccnwsngggncccwsnwsnmgngcngtnmgn 1383
           |:||
Db   1508  gytaygaraararytnaaacycngargacngacnaarggna-----aytgyggnaargayg 1562

Qy   1384  mgnwsncccccnwswsnmgnytncararggnmgnwsnacngaywsnytncarcaygtn 1443
           |:||
Db   1563  ngaymgntggathcartggwsnaarcaygaygtntttgyggnttytntgtgyacnaa 1622

Qy   1444  ccngararwsnacngayacncaartgycarccngtnaargcngcnggnaaytgygggnaagtn 1503
           |:||
Db   1623  yytnacnmgncccmgnathggncarytnacaaracngtnaggnarathathccnacnwsnttyta 1682

Qy   1504  ccntayaaracngtngcngaryytnacaaracngtnggngnathtayytyntncaytgy 1563
           |:||:
Db   1683  ycaycargnmgntnathgaytgywsnggngcncaygtngtyntngaygaygayacnga 1742

Qy   1564  caygayytngaytngcaygggntnaaamgngaycaytyggngcnytnmgnttygay 1623
           |:||          ::|
Db   1743  ygtnggntaygtngargaggnacncctggyggnccnatgatgcgyytngaygnaa 1802

Qy   1624  tgyccnacngnttymgnacntayatgggncnccngtnaccnytntgyttgyacgca 1683
           |:||:
Db   1803  rtgyytncaratncargcnytnaaayatgwsnwsnytgyccnytngaywsnaargnaargt 1862

Qy   1684  ccnttygnacncgntgtttyacncartgyytnytayytncayt 1726
```

```
Db    1863  ntggywsnggncayggngtntgywsnaaygaargcnacntgyath 1905
            :::   |       |      |::|   |::|  |  | :|| : :|::

RESULT 14
US-09-007-005-17
Sequence 17, Application US/09007005B
Patent No. 6258558
; GENERAL INFORMATION:
;   APPLICANT: Szostak, Jack W.
;   APPLICANT: Roberts, Richard W.
;   APPLICANT: Liu, Rihe
;   TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;   TITLE OF INVENTION: FUSIONS
;   FILE REFERENCE: 00786/350003
;   CURRENT APPLICATION NUMBER: US/09/007,005B
;   CURRENT FILING DATE: 1998-01-14
;   EARLIER APPLICATION NUMBER: 60/035,963
;   EARLIER FILING DATE: 1997-01-27
;   EARLIER APPLICATION NUMBER: 60/064,491
;   EARLIER FILING DATE: 1997-11-06
;   NUMBER OF SEQ ID NOS: 33
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match             2.8%;  Score 48.6;  DB 4;  Length 289;
Best Local Similarity   10.1%;  Pred. No. 0.00038;
Matches  27;  Conservative 101;  Mismatches 139;  Indels   0;  Gaps  0;

Qy   743  argargarathwsnaarcarcaarwsnathcargargtnacntgggtnytnaytnaargcnt 802
          :::  | :::::|||:: :|||:|| |::|| |:|::| ||:|| ||:||
Db    19  rarcrurarurururarcrararurarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy   803  tywsnttyathmgnqargcngarcayaaarwsnwsnqaraayytncayccnqayaayytna 862
          : ::|:::|| |:|||:|:|:|:::||||::::|:|::::| |:: |::|: ::||  :::
Db    79  rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy   863  thaaraaaraaayccnttywsnqarggnaarttyaarytngcngcnqarathtgyatht 922
          ::: |  | :|::|  | |::|  ::|    ::|  :::
Db   139  rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy   923  gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 982
          :::  :::|     ||||:|:::|  | |::|:: :  |::|:::::  :   |
Db   199  rsrnrnrsrnrnrsrnrnrscrargrcrurgrcrurarcrurcrururgrcrgrc 258

Qy   983  armgnwsnwsncarcarwsnathaarw 1009
          : ::|:::| |:::|   :  :|:::
Db   259  ruaaaaaaaaaaaaaaaaaaaaaaaaa 285


RESULT 15
US-09-244-796-17
Sequence 17, Application US/0924796
Patent No. 6281344
; GENERAL INFORMATION:
;   APPLICANT: Szostak, Jack W.
;   APPLICANT: Roberts, Richard W.
;   APPLICANT: Liu, Rihe
;   TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;   TITLE OF INVENTION: FUSIONS
;   FILE REFERENCE: 00786/350007
;   CURRENT APPLICATION NUMBER: US/09/244,796
```

```
;   CURRENT FILING DATE: 1999-02-05
;   EARLIER APPLICATION NUMBER: 60/035,963
;   EARLIER FILING DATE: 1997-01-27
;   EARLIER APPLICATION NUMBER: 60/064,491
;   EARLIER FILING DATE: 1997-11-06
;   EARLIER APPLICATION NUMBER: 09/007,005
;   EARLIER FILING DATE: 1998-01-14
;   NUMBER OF SEQ ID NOS: 33
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match             2.8%;  Score 48.6;  DB 4;  Length 289;
Best Local Similarity   10.1%;  Pred. No. 0.00038;
Matches  27;  Conservative 101;  Mismatches 139;  Indels   0;  Gaps  0;

Qy   743  argargarathwsnaarcarcaarwsnathcargargtnacntgggtnytnaytnaargcnt 802
          :::  | :::::|||:: :|||:|| |::|| |:|::| ||:|| ||:||
Db    19  rarcrurarurururarcrararurarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy   803  tywsnttyathmgnqargcngarcayaaarwsnwsnqaraayytncayccnqayaayytna 862
          : ::|:::|| |:|||:|:|:|:::||||::::|:|::::| |:: |::|: ::||  :::
Db    79  rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy   863  thaaraaaraaayccnttywsnqarggnaarttyaarytngcngcnqarathtgyatht 922
          ::: |  | :|::|  | |::|  ::|    ::|  :::
Db   139  rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy   923  gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 982
          :::  :::|     ||||:|:::|  | |::|:: :  |::|:::::  :   |
Db   199  rsrnrnrsrnrnrsrnrnrscrargrcrurgrcrurarcrurcrururgrcrgrc 258

Qy   983  armgnwsnwsncarcarwsnathaarw 1009
          : ::|:::| |:::|   :  :|:::
Db   259  ruaaaaaaaaaaaaaaaaaaaaaaaaa 285


Search completed: July  8, 2002, 16:22:51
Job time: 12295 sec
```

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: July 8, 2002, 18:50:56 ; Search time 3393.27 Seconds
(without alignments)
6885.167 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargntaygg..........gyytntayytncaytcygatg 1731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	479.2	27.7	2615	10	BE420422	BE420422 32-393 hu
	2	450.2	26.0	2009	11	BC004496	BC004496 Homo sapi
	3	398	23.0	1514	10	BE512633	BE512633 32-1514 h
	4	336.8	19.5	870	10	BI488505	BI488505 603021014
	5	335.6	19.4	1050	10	BM472108	BM472108 603020179
	6	321.8	18.6	1036	10	BM471183	BM471183 AGENCOURT
	7	318.0	18.4	898	10	BM457166	BM457166 AGENCOURT
	8	313	18.1	717	12	AG086951	AG086951 Pan trogl
	9	306.0	17.7	736	10	BI754555	BI754555 603023384
c	10	300.4	17.4	654	12	AG112666	AG112666 Pan trogl
	11	296	17.1	676	12	AG064424	AG064424 Pan trogl
	12	287.0	16.6	691	12	AG062447	AG062447 Pan trogl
	13	286.2	16.5	720	12	AG116938	AG116938 Pan trogl
	14	283.2	16.4	677	12	AQ037711	AQ037711 CIT-HSP-2
	15	283.0	16.4	677	12	AG097258	AG097258 Pan trogl
	16	282.4	16.3	704	12	AQ588828	AQ588828 Pan trogl
	17	281.6	16.3	659	12	AG092858	AG092858 Pan trogl
	18	281.4	16.3	701	12	AQ588873	AQ588873 CITBI-E1-
	19	281	16.2	729	10	BM045873	BM045873 603624875
	20	279.6	16.2	760	10	BG281182	BG281182 602402107
	21	279.4	16.1	697	12	AG018926	AG018926 Homo sapi
	22	278.8	16.1	885	10	BG761979	BG761979 602718930
	23	277.2	16.0	1182	10	BM451317	BM451317 AGENCOURT
	24	276.2	16.0	639	12	AG115932	AG115932 Pan trogl
	25	275.2	15.9	676	12	AG125093	AG125093 Pan trogl
	26	275	15.9	721	10	BE728616	BE728616 601563027
	27	274.8	15.9	671	12	AG045131	AG045131 Pan trogl
	28	274.6	15.9	877	10	BI916691	BI916691 603178655
	29	272.2	15.7	625	12	AG070563	AG070563 Pan trogl
	30	271.4	15.7	704	12	AQ020199	AQ020199 CIT-HSP-2
	31	271.2	15.7	686	12	AG119684	AG119684 Pan trogl
	32	268.8	15.5	660	12	AG057341	AG057341 Pan trogl
	33	268.6	15.5	746	10	AG001408	AG001408 Homo sapi
c	34	266	15.4	623	12	AG075663	AG075663 Pan trogl
	35	266	15.4	699	9	AL602279	AL602279 DKFZp686B
	36	264.8	15.3	543	12	B94403	B94403 CIT-HSP-217
	37	263.8	15.2	675	12	AG096655	AG096655 Pan trogl
	38	261.4	15.1	833	9	AU140358	AU140358 Pan trogl
	39	261	15.1	771	12	AG001229	AG001229 Homo sapi
	40	260.6	15.1	716	12	AG104336	AG104336 Pan trogl
c	41	260.6	15.1	940	10	BG679573	BG679573 602627616
	42	260.6	15.1	955	10	BG681769	BG681769 602627877
	43	257.4	14.9	649	9	AW857140	AW857140 RC1-CT030
c	44	256.6	14.8	900	10	BM449699	BM449699 AGENCOURT
	45	255.4	14.8	676	12	AQ239408	AQ239408 CIT-HSP-2

ALIGNMENTS

RESULT 1
BE420422

LOCUS   BE420422   2615 bp   mRNA   linear   EST 11-OCT-2001
DEFINITION 32-393 human bone marrow cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BE420422
VERSION   BE420422.1 GI:16041640
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2615)
  AUTHORS   Lu,X.; Cui,L. and Li,Y.
  TITLE   DDRT-PCR from B cell
  JOURNAL   Unpublished (2000)
COMMENT   Contact: xingwu Lu,liangxian Cui,yonghai Li
            Department of Biochemistry
            Institute of Basic Medical Science, Peking Union Medical College
            DongDan SanTiao 5, Beijing, P.R.C, 100005
            Tel: 86-010-65296951
            Email: luxingwu@263.net
FEATURES   Location/Qualifiers
     source   1..2615
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="human bone marrow cDNA library"
                     /tissue_type="bone marrow"
BASE COUNT   705 a   531 c   641 g   738 t
ORIGIN

Query Match 27.7%; Score 479.2; DB 10; Length 2615;
Best Local Similarity 47.1%; Pred. No. 2.6e-99;
Matches 500; Conservative 222; Mismatches 299; Indels 41; Gaps 4;

Qy   706 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 765
          ||:|:|| |:| ||| || | |: |||::||::|||::||||
Db   494 AACCTTGAGCTTGAGAGAGATCATTTAGGGTATCTGGCAGGAGAGAAATTTCTAAGCAGCA 553

```
Qy    766 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmngargcngar 825
          :: !.:||:|:||..:| |.::|||..||:| ||||..|.||: ||:||:||.|:..|
Db    554 AGCATTCAAGAGGTAACTTGGGTGCTGTTAAAGGCATTTAGTTTTATAAGGGAAGCAGG 613

Qy    826 cayaayarwsnwsngaaraayytyncaycccngayaaygtnathaataaraaraayycnttywsn 885
          :||:|:||:::|:|: ||| |.::||||::..::||.:|.:|:|: |:..:.||:::
Db    614 CATGAAAGTTTGGAAAATTTGCAGCACAATCCAAATGAAAAAGAAAAATCCCATATTCT 673

Qy    886 garggnaarttyaarytngcngcngaarathtgyathtgyaaygargaryytnaaygtnaay 945
          |||.:::|:.||:: ||||:|||:|::|.|:::|:||.:::.||:||.|::||||.|.:|
Db    674 GAGGGAGAAAATCAAGCTGCCTGCGCAGAAAATTTGCATATGTAACAAGGAGTTAAAGTGTAAT 733

Qy    946 cncncargayaayggngaraayathwsntggactgycyarmgnwsnwsncarcarwsnath 1005
          || |::||||:|.||:.|:::|::|.::||.|:|.|:::|||..::||:||.||||:|::
Db    734 CCCCAAGACAAATGGGAAAATGNTCCAGGACATGCACAAGTGCTTCACAGCAGCCCTC 793

Qy   1006 aarwsnytycntggmgnccnmgnmg----naattggttytgtyggnacnggncnggnwsn 1062
          ||:|::|:::.::::.||::|::| :.:::|||.:||:|.|:.|||.|||.:||.||.
Db    794 CCATCACAGGCCTCGAGGAGAAAAAGTCATCTCTCTGGGCCAGGCCCAGGGTCC 853

Qy   1063 ytntggytngtytncarccnmgngayytngtnccntgggtnccngtnaaywsngc------ 1115
          |||.::.|::|::::|||.::||.:||.||||.||::.|||:||||.:||||
Db    854 CCAGGCTGTGTGCAGCCTGGGACTTGGTGCCTTGTGTCCCTGTTGTCCCAGCCTGCCT 913

Qy   1116 ------------ngtngcnwsngargngngcnwsnccnaarccntgg 1149
                      |.||| :| ||.||.||.||:|:|:|..:.|::||||.
Db    914 GAAAAGGGGCCCAACATAGAGACTCGGATTGTCGGCTTCACAGGGTGCAAGCGTGTAACCTTGG 973

Qy   1150 carytncnwsngngtngarccngtngngcnaaraatwsnmgnathgaargtntgggar 1209
          ||:|::|:||||:||||||::||::||.||.||:|::.|::||.|:::|:|:|||||
Db    974 CAGCTTCCATGTGTTGAGCCTGCCAGTGCACAGAAGTCAAAAATTGGGGTTTGGGAA 1033

Qy   1210 cncccnathmgnttycaraarathtayggnaayccntggatgccnmgncaaraattygcn 1269
          || |::|:|.||.||:||||||.|..|||..|||:||.:|||.||:|||||||:|||:|
Db   1034 CCCTCTGCCTAGATTTCAGAAGATGTATGGAAAATGCCCTGGCAGTCGTGCTGCT 1093

Qy   1270 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraaraargngnaaygtnggn 1329
          ||.||||||.||.||:||||.||:|.:|.|||:||||.|::||||.|::|||:.||:|||.
Db   1094 GTAGGGAGGGGCCCTCATGGAGAATCTCTGCTACAGAAGGGAAAATGTGGGA 1153

Qy   1330 tgggarcacnccncaaymgngtnccnwsnwsnmgngcngcnaacwsncngtnmgngngwsn 1389
          ||||||.||.||:|::|:|:|||||.||.||.||.||:||::||.::|:|||:|||||.||
Db   1154 TCGGACCCCAGACACAGAATCCTACTGGGGCACCGCCTAGTGGAGCTGTGAGAAGAGGG 1213

Qy   1390 cncccnwsnwsnmgnytncaraarggnngnwsnacngaywsnytncarcaygtnccngar 1449
          || ||:||.||.||.||:||||| ||.||:|::||.||.:||:|||:.:||.||.|||.:
Db   1214 CCACCATCTCCAGACCCGAAATGATAGATCCACTGACAGCGCTGCTCACTGGA 1273

Qy   1450 aarwsnacngayacncaartgycarccngtnaargcngcnggnatggaarwsngtnccntay 1509
          ||:||:||:.||.|:||:||||:|:|||::||.||.|||:|:||||.||.||:.||:||:|
Db   1274 AAAACCGCAGA----CAACACACGCCATGAAACGCACTAGGCACGAGGCGTCACCTGC 1329

Qy   1510 aaracngtngtnycacnccagcngaryytnacnaaracngtngcnaanaathtayytnytnytncaycaygtgycaygay 1569
          ||:||:||||:|:|:|::|.:|||:::||||:|::|||:|||::||.|..|:::|||.::|||:|::|.|.|:||||||
Db   1330 AAAGCCACACAGGGCGCGGAGCTGCCCAAGATCATGGAACCCACCTCTTGCATCAGCATGAC 1389

Qy   1570 ytnaaygtnmgncaygggngtnaarmgnaaycaytymgngacnytnmgnttygaytgyccn 1629
          |::|||||:|||.|.||.||:||:.|||.|:||||.:|::||.|::|||::||||:||:||
Db   1390 CTGGAGTGGAGATTTAAGGGAAGCAGCGGCATAAAAGTTCAGAAAATTTGACTGCCC 1449

Qy   1630 acnggnttymgnacntayatgggngccngt-ncnytntgytyggncaarttyttyccntt 1688
          ||.||:||::||:||.|:|||||.|||:|| .|:.|:|:|:|||:||:|||:|:|:|:||
Db   1450 TCTAGATTTCAGACTTGCATGGGCCCTGTAGCCCTTTGTTTTACTCAATTTCTCCATT 1509

Qy   1689 yggnacngcngtntntyacncaartgyytncaytyat 1730
          ||||||:|||||:|:|:||:|||:||.:||:||||:|:|
Db   1510 TGGAACAGCTGTGTTTAACACAATACCTGTACCCCCATTGTAT 1551
```

```
Query Match  26.0%;  Score 450.2;  DB 11;  Length 2009;
Best Local Similarity  45.2%;  Pred. No. 1.1e-92;
Matches   452;  Conservative  214;  Mismatches  321;  Indels  13;  Gaps  3;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          |:||||||:|::||.|||.||.||:|::|||||.||.|.|.||||.|:|::|:|||||:|
Db    489 AGAAGAAATTTCTATGCAGCAAAGCATTCAAGAGTGTGACTTGGGTGCTGTTAAAGGCATT 548

Qy    804 ywsnttyathmngargcngarcaayaarwsnwsngaraaaytyncaycccngayaaygtnat 863
          |.:|::|||.|||.||.||.|||.|:||||||.||.||:.|::|||.||:.||.||:|||
Db    549 CAGATTTATAAGGGAAGCAGCGGCATAAAAGTTCAGAAAATTTGACTGCCC 608

Qy    864 haaraaraaraayycnttywsngargngnaarttyaaarytngcngcngaarathtgyathtg 923
          :|||||||||.||:||||||:||:||.||.|||:|||.|:.||:|||:|::|.|:|||.|::
Db    609 AGAGAAGAAAAACCCCATTTTTTGGAGAGAAAATTCAACCCAGCTGCATAAAATTTTGCATAAG 668

Qy    924 yaaygargaryytnaaygtnccncargayaaygtnccagcngaaaatgggnaaattywsngar 983
          |.:|:|||||:::||||||.||:.||:|||.||:.|.||:.|||.|.|||||.:|:||||.||
Db    669 TAACAAGGAGTTAAAGTGTAATCCCCAAGACAAATGGGAAAATGNTCCAGGACATGCACAAG 728

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytycntgmgnccnmgnmgnaatggttytgtyg 1043
          ::|:.::|::|||.||:|::||:|..||:||:||:||.||:|.||.|:|||.:||:|.|::|
Db    729 GAGGTCTTCACAGCAGCCCTCCCGTCACAGGCCTCGAGGAGAAAAAGTCATCTCTCTGGG 788

Qy   1044 yggnacngcngncnggnwsnsnytnt------gytgtgytncarccnmngngayytngtgtncc--- 1094
          ||||.|||.|::|.||.||.:||.|:       |:.|:|:|:::||.::|:|||::||:|:.||:
Db    788 GCCAGGCCCAGGGTCCCAGGCTGTGTGCAGCCTGGGACTTGGTGCCTTGTGTCCCTGTTGTCC-- 1094
```

```
FEATURES
         source          Location/Qualifiers
                         1..1514

COMMENT         Contact: xingwu Lu,liangxian Cui,yonghai Li
                Department of Biochemistry
                Institute of Basic Medical Science, Peking Union Medical College
                DongDan Santiao 5, Beijing, P.R.C. 100005
                Tel: 86-010-65296951
                Email: luxingwu@263.net
                full cDNA sequence.

JOURNAL         Unpublished (2000)
TITLE           DDRT-PCR from B cell
AUTHORS         Lu,X. Cui,L. and Li,Y.
REFERENCE       1 (bases 1 to 1514)
DEFINITION      32-1514 human B lymphocyte cDNA library Homo saneins cDNA, mRNA
                sequence.
LOCUS           BE512633        1514 bp        mRNA        linear        EST 11-OCT-2001
RESULT  3

BE512633
```

```
ORIGIN
BASE COUNT        380 a        341 c        398 g        395 t

        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="human B lymphocyte cDNA library"
        /tissue_type="human bone marrow"

Query Match        23.0%;  Score 398;  DB 10;  Length 1514;
Best Local Similarity   47.1%;  Pred. No. 1.1e-80;
Matches  427;  Conservative  182;  Mismatches  256;  Indels  41;  Gaps  4;
```

(Sequence alignment data — rows labeled Qy (Query) and Db (Database) with base positions and alignment bars)

```
Qy   1512 ...
Db   789  TGGACCTGTGCCAGCCTAGGTGTTACTGGCACCCTGTGTCCCAGCTACTCCAGTGTGGCCTGA  848
...
```

```
RESULT  4

Db    897 TTGTAT 902
          ::|:||:||

LOCUS       BI488505
DEFINITION  BI488505
            6030210I4F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191803 5',
            mRNA sequence.
                              870 bp    mRNA    linear    EST 28-AUG-2001
ACCESSION   BI488505
VERSION     BI488505.1  GI:15327733
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1479 row: k column: 04
            High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1..870
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5191803"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning). Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      222 a    211 c    256 g    181 t
ORIGIN

Query Match          19.5%;
Best Local Similarity  44.6%;    Score 336.8; DB 10; Length 870;
Matches   378; Conservative 161; Mismatches 273; Indels  35; Gaps   3;
                                 Pred. No. 1e-66;

Qy    847 caycccngayaaygtnathaaraararaayccnttywsngarggnaarttyaaarytngcn 906
          |||:||:||:|:||:||:||:|:||:||::||::|||::||:|::||::|::|||||:|:
Db      2 CAGCCTGATCATGAGTAGAGAAAAGAAAAACCCATTTTTTGAGGATAAATTCAAGCCAGCT 61

Qy    907 gcngararathtgyathtgyaaysyargaryytnaaygtnaaaygtgayaaygngngaraay 966
          ||:|||||:|:||:||||:||:||:||:||:||:||:||:||::|::|:||:||||||||
Db     62 GCAGAAAATTTGCATATNGTAGCAAGAGAGTCTAATGTTAATCTCCAAGACCATNGGGAAAAT 121

Qy    967 athwsnttggacntgyaarmgnwsnwsncacaarwsnathaarwsnytngcntggmgnc-c 1025
          |:| |:||||| || || ||| |||::| ::| |||::||:||:||:| || ||||| |
Db    122 GTCTCCGGGCCCATGGCAGAGACCTTCATGGCAGCCCTCCCATCACAGGCTGGAGCCAG 181

Qy   1026 nmgnmgnaartggttytgyggnacnggncacnggncatccngctg-ngtncarccnmnga 1085
          ::| ::|:| ||:|::| |: :|||:||||::| ||||:|||:||:||||||:|||:|||
Db    182 GAGGAAAAAGAGGTTTCGTGGGCCAGCCATCCCGTGCTTTGTGCAGCCTAGGGA 241

Qy   1086 yytngtnccntgygtnccngt---------------------------naayws 1112
          ::|| || || |||:|| || ||                         :| :
```

```
RESULT  5

Db    242 CTTGGTGCCCGTGTCCCAGTGCGCTCCAGCTGTGACTGAAAGGGGCCAAGCTAGAGCTCG 301
          ||:|||||: |||||||||:|:|||:||||:|||:|||||::|::||||||||||||:||
Qy   1113 ngcngtngcnwsngargngwsnccnaarccntggcarytnccnwsnggngtngarcc 1172
          ||:||:|||:||::|::||:|||:::|:||:|:::||:||::|::|:||::||:||
Db    302 GGCTGTGGGCCTTCAGAGGGTGGAAGCTCTAAGCTTTGGCAGCTTCCACGTGGTGTTGAGCC 361

Qy   1173 ngtngngngcnaaaraarwsnmgnathgargtntgggarccnccnacathmgnttycaraarat 1232
          ||:|:::|:|||:||||:||:||:||||||:||::|:||||||:||||::||||:||||||||
Db    362 TACAGGTGCACAGAGAGTCAAGAAACTGCAGGTTTTGGGAACCCTGCCTATGATTTCAGAAGAT 421

Qy   1233 htayggnaaycccntggatgccnmgncaraarttygcngtngngtnggnnwsnwsnntggng 1292
          :||:||::||:|||||||||||::||:||||::||:||||:||:|::|:||::||::|||::
Db    422 GTTTTGGAAATGCCTGGATGCCCAGGCAAAAAGTTTGCTGCAGGAGTGGAGTGGGCCTTATGGAG 481

Qy   1293 nacnwsngcnnmgngtngtncaaraargnaaygtngngttgggagrccncncncaymgngtncc 1352
          |||:||:||:||:||:||:||||:|||:||||:||:|:||:||||||::||:|:|||:||:||
Db    482 AACCTCTGCTGGAGCAGTGCAGAAGGAAAAATGTGGGATCGGGACAGAGAGTCCA 541

Qy   1353 nwsnggngcnccnwsnwsngngcntnmgnmgnwsnccncccnwsnwsnmgngnytncaraa 1412
          :||:||::|:|||::|:||||:||:||:||:||:||:||:|:|::||::|||::||||||
Db    542 TACTGGGCACTGCCTAGTGAAACTGTGAGAGAACACCACCGTGCTCTCCAGAGCCCAGAA 601

Qy   1413 rggnmgnwsnacngaywsnytncarcaygtnccngararaarwsnacngayacncaartgyca 1472
          ||||:|||::|||:||::||:||||:||:|:||:|||:||||:||||:|::|::||:||:||
Db    602 TGGTAGATCACCAACAGCTTGCACCATGTGCCTGAAAAAGTCAGGTACTCAATGCCA 661

Qy   1473 rccngtnaaargncngcnggnaggarwsngtnccntayaaracngtngtncgngarytnac 1532
          :||:||:|||:|:||:||::|:||||:||:||:|:||||||:||:||:||:||||:||:||
Db    662 GCCCATGAAGGCAGCCAGGAGGAGGCTGTACCCTGCAAAAGCCACACAGGGAGGGAGCTGTC 721

Qy   1533 naaracngtnggnaathtayytnytncaygtycaygayytngaygtnagncaygggngtnaa 1592
          :|||||:||:||::|:||::||:||:||:||:||:||::||:||:||:||:|:||::||:|
Db    722 CAAGACCATGGGAAACCACCTTGACCCTTGGATGCGAGACCTTGGAGTCAA 781

Qy   1593 rmgngaycayttygngncnyt-nmgnttygaytgtyccnacnngngnttymgnacntayatgg 1651
          :||:||:||:||:||:||||| |:||:||:||:|:||:||||:||:||:||:|||:||||||
Db    782 AGGAGGATCATTTTGGGACTTTAACAAATGTGACCTGCCTGCTGGATTTCAGACTTGCATTG 841

Qy   1652 gnccngt 1658
          | || |
Db    842 GGGCCCT 848

LOCUS       BM472108
DEFINITION  BM472108
            AGENCOURT_6465359 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5539381
            5', mRNA sequence.
                              1050 bp    mRNA    linear    EST 05-FEB-2002
ACCESSION   BM472108
VERSION     BM472108.1  GI:18521150
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1050)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12233 row: m column: 14
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1..1050
                     /organism="Homo sapiens"
```

RESULT 6
LOCUS   BM471183   1036 bp   mRNA   linear   EST 05-FEB-2002

Left column

```
                /db_xref="taxon:9606"
                /clone="IMAGE:5563381"
                /clone_lib="NIH_MGC_72"
                /tissue_type="melanotic melanoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: skin; Vector: PCMV-SPORT6; Primer: Oligo dT.
                Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Library constructed by Life
                Technologies."
BASE COUNT      295 a    262 c    272 g    221 t
ORIGIN

Query Match     19.4%; Score 335.6; DB 10; Length 1050;
Best Local Similarity 43.8%; Pred. No. 2.2e-66; Matches 335; Conservative 169; Mismatches 248; Indels 12; Gaps 2;

Qy   744 rgargarathwsnaarcarwsnathceargargtnacntgggtnytnaaygtnaat  803
         :|:.|:||||.||.:||| ||.||.:||:||| ||.||||||||||||||:|||||.||
Db   ...

Qy   804 ywsnttyathmgnargcngargcarcyaearwsnwsngaraayytncayccngayaaygtnat  863
         |.:|||| |||||||||||||||||.||||:||||||||:||||||||||||.|||||:||
Db    16 AGAAGAAATTTCTATGCAGCAAAAGCATTCAAGAGTTGACTTGGGTCGTCGTTAAAGGCATT  75

Qy   864 haaraaaraataayccnttywsngargngnaaattyaarytngcngaratthtgyathtg  923
         :||||:|||||||||.:|||||||||||||:||||||||||||||:|:||:|||:.||||
Db    76 CAGAGTTATAAGGAAGCAGAGCAGCCATTTTTTGAGGAGAAATTCAAGAAAATTTGCATAAG  135

Qy   924 yaaygargarytnaaygtnaayccncargayayayggngaraayathwsntggacntgyca  983
         :||:||||||||:|||||||:|||||:|||||||:|||||||||.||||||||||||||||
Db   136 AGAGAGAGAAAAGCCCATTTTTGCAGAGACCATAAAAAGTTCACTGCAGAAATTTGCATAAG  195

Qy   984 rmgnwswsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaattggttytg  1043
         ||:||||||||||||:|||||:||||||:||||||||.|||||||||||:||:.|||:|
Db   196 TAACAGAAAGCCGCATGTTAATCCCAGAACACACAGGAGAGACATCTCCAGGACATATCA  255

Qy  1044 yggnacngngnccnggnvsnytnt-----gytgtgtncarccnmgngayytngtncc---  1094
         :||||||||:|||||||.:||||.       ||.||||||||||||:||||.|||||
Db   256 GAGGTCTTCACCACCAGCCCTCCCGTCACCAGGCCCTAGGAGAGAAAAATGGTTTCC  315

Qy  1095 ---ntgygtnccngtnaaywsngcngtngcnwsnsgargaargaryt--naaygtnaaycc  ...
            ||||||||||||||||||.:||||||.:|||:||||||||||||   |||||||||||
Db   316 TGGACCTGTGCAGCCTAGGTACTTGGCACCCTGTCCCAGCTACTCCAGTAGTGGGGTGA  375

Qy  1152 rytnccnwsngngtngarccngtnggngcnaarraarwsnmnathgargtntgggarcc  1211
         ||||||:|||||||||||:||:|||||||||||||:||||||||||||||||||||||
Db   376 AAGGGGCCCAAAGTACAGCTTACCGCCATGGCTTCAGAGGGTGCAAGCGCCTTGGCA  435

Qy  1212 ncnnathmgnttycaraaarathtayggnaayccntggatgcgcnmgncarccncacncnm  1271
         |||||||:|||:|||||||||||||.:|||||||||:|||||||||:|.||:||:||:
Db   436 GCTTCCCCGCATGCTGTTGAGCCCTGCAAATGCACAGAGAATTTGGATTTGGGAACC  495

Qy  1272 ngggtnggnwsnsntggmgnacnwsngtngtngcaragnnmgnwsnaccrtnggggntg  1331
         |||||||||||||||||||||||.:|||||||||||:||:||:||||:||.||||||
Db   496 TCTGCCTAGATTTCAGGAGATGTACCGGAAGTGTCCGGATGGGAACATGGGGTC  555

Qy  1332 ggarccnccncaymgnytnccnwsnsngggnaacnwsnsngngcnhcnwsnmgnwsncc  1391
         :|||:|||||||||||:|||||||||||||||:|:|||||||||||||:||||:|||:
Db   556 AGGGGCAGGGACGGCCTTCATGGAACCTCTGCTACGGAAGATGTACCGGAAGTGTCCC  615

Qy  1392 ncnwsnwsmgnytncaraargghmgnwsnacngaywsnytncca  1451
         |||||||||:|||||||||||:||||:||||||||||:|||:|
Db   616 TGAGCCCCCACACAGAATCCCTACTGAGGCGCATTCAAGAGAAA  675

Qy  1452 rwsnacngayacncartgycacccngtnaargcngcngnnatgg  ...
         :|||||||:||:|||||||||:||:|||||||||||||||:|||
Db   676 ACCATCCTTCCACACACCCAGAAATGATAGATCCACACAGCAACACGCTCGACCACCCTGGAA  735

Qy       ... 779
Db   736 AGCCCGCAGAGCATTCAACACCAGCCCGGGGAGCCAGCCAAAAGGG  779
```

Right column

```
DEFINITION  AGENCOURT_6478314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5563049
            5', mRNA sequence.
ACCESSION   BM471183
VERSION     BM471183.1  GI:18520225
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: cgapbs-r@mail.nih.gov
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            Tissue Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Library Prepared by: Agencourt Bioscience Corporation
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1293  row: g  column: 18
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1..1036
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5563049"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: PCMV-SPORT6; Primer: Oligo dT.
                     Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Library constructed by Life
                     Technologies."
BASE COUNT      254 a    239 c    294 g    247 t  2 others
ORIGIN

Query Match     18.6%; Score 321.8; DB 10; Length 1036;
Best Local Similarity 45.0%; Pred. No. 3.4e-63; Matches 390; Conservative 166; Mismatches 267; Indels 43; Gaps 6;

Qy   907 gcngaratthtgyathtgyaaygargaryt--naaygtnaayccncacargayaaygngnara  964
         |||||||||||:||||||:||||||||||    |||||||:|||||||:|||||.||||||:
Db     1 GCAGAAAAATTACATATAAGTAGCACAACTACAATAGTTTAATCCCAAGACACCATGGGGAAA  60

Qy   965 ayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngtytgygtgtca  1024
         ||||||||||||||||||||:||||||||||||||||||||||:||||||||:||||:||||
Db    61 ATGTCTCCAGTTCCTGTCAGGAGGACCTTCACTCAGCAGCCCACCAGAGC  120

Qy  1025 cnmgnmg---naartggttgyggnaacnggncenggnwsnsnytngtytgygtgtncarccnm  1081
         |||||:|   ||||||||:|||||||||||||||||||||||||||:||||:||||||||:|
Db   121 CCAGGAGGAGAAAAAGTGGTTTCGTGGGCCTGGGTTCAGGGTCCCCATGCTGTGTGAAGCCTA  180

Qy  1082 gngayytngtncnctgygtnccng-----------tna  1108
         |||||||||||:||.||||||||            |||
Db   181 GGGACTTGGGTGCCCTGTGTCCCACCTGCTTCCCACTGGCAA---ytnccnwsngggng  240

Qy  1109 aywsngcngtngcnwsnsngargngngcnaaraarwsnmgnathgargtntgggarccncca  1166
         ||||||||||||||||||||:||||:|||||||||||||||||:||||||||||||||||:|
Db   241 CTCTGCGTGGCTGGCTTTAGAGGGTGCAAGCGCCTTGGCAGCTTCTTCCATGTGGTG  300

Qy  1167 tngarccngtnggngcnaaraaarwsnmgnathgargtntgggarccncca  1225
         :|||||||||||||||||||||:|||||||:|||||||||||||||||||:
Db   301 TTGAGCCTGTGGGTGCACAGAAGTCAAGAACTGAGGTTTGGGAACCTGCT-AGATTC  359

Qy  1226 araaratthtayggnaayccntggatgccnmgncararaarttygcngtnggngtnggngnw  1285
         ||||||||:||||||||||||||||||||:|.||:||||||||:|||||||||||||||||
Db   360 AGAAGAATGTATGGAAAATGCCTGGATGCCCCAGCCAAAAGTTTGCTCGTGTGAAGGGCCT  419

Qy  ... 1286 sntggmgnacnwsngcngmgngtngtncararggnaaygtnggngtnggarccncncaym  ...
Db   ...
```

```
Db   420 CATGGAGAAACCTCTGCTAGGCACTGTGGAAGGAAAATGTGGGTTGGAGCCCCGCACA 479
         ||.||  |||| || .||| :|| .|| ||||||  ||  ||||||| || || .||::
Qy  1346 gngtnccnwsnggngcnccnwsnwsnmgncngtnmgngnmgnwsnccnwsnwsnmgny 1405

Db   480 GTGCCCCTGCTGGGCACTTGGCACTGCTGCCTAGTGCAGTGAGAAGAGGGCCACATCCTCCAGAC 539
         || || .|| |||| || ||.||.|| .:| || ||.|||||| || || .||||| || ||
Qy  1406 tncaraaargggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngaywsnccnga 1465

Db   540 CCCAGAAATGGTAGATCCATCAACACGCTTGCACTGTGCCACTTGGAAAAACCCGACACACTC 599
         .||| |||| ||  :|| ||.||.||.|| .|| ||.|| .:| ||||| |||| ||.||.:|
Qy  1466 artgycarccngtnaargcngcngcnggnatggarwsngtnccntayaaracngtngtngcng 1525

Db   600 GACAACCAGCCCATGAAAGCAGCTGGGAGGGAGGGCTATACCCTGCNAAAGACACAGGGCAG 659
         .||.|| ||| ||| ||.||.|| ||.||.||.||.|| ||.|||||.||||||||||||||
Qy  1526 arytnacnacngtngngnaathtayytnaytcaytgycaygayytngaygtnmgncaryg 1585

Db   660 AGCTGCCTAAGCACGATGGGGAACCCACCTCTTCGCATCAGCANTGACCTTGGATATGAGACTTG 719
         .||.|| |||.|| |||||| |||| || ||.||.||.|| ||.|| ||.|||||.||.||.||
Qy  1586 gngtnaaaringngaycayttyggngcnytnmgnttygaytgyccnacngnttymgnacnt 1645

Db   720 GAGTCAAAAGGAGAATCANTTTTGAGCTTGAAAATTGATTGCCCGCGTGGATTTCAGACTT 779
         .||.|| |||.|||| ||.||.|| ||.|| |||| ||.|||||||.|| |||||||| ||
Qy  1646 ayatgggnaccngt-nccnytntgyttyggncaartytyccntttgggnacngcngtntty 1704

Db   780 GTGTNANGCCTGATNAACCCCTTTGTTTTTGGGCAATTCTCCCATTTTGAATGGCTACANTTT 839
         .||.||.||| ||.|||||||||||||||||||||||.|||||.||||||||||||||.|||
Qy  1705 acncartgyytntatytncaytgyat 1730

Db   840 TACCAATACCTGTACCCCTATGGGAT 865
         ||||||||||||||||
```

RESULT 7
LOCUS   BM457166   898 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427
            5', mRNA sequence.
ACCESSION   BM457166
VERSION   BM457166.1 GI:18506206
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12346 row: h column: 20
            High quality sequence start: 9
            High quality sequence stop: 713.
FEATURES   Location/Qualifiers
     source   1..898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5583427"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

BASE COUNT   225 a   228 c   254 g   191 t
ORIGIN

Query Match   18.4%; Score 318.6; DB 10; Length 898;
Best Local Similarity   45.1%; Pred. No. 1.7e-62;
Matches   389; Conservative 162; Mismatches 253; Indels 59; Gaps 5;

```
Qy   916 tgyathtgyaaygargarytnaaygtnaaytgyccnacargayaaygggnagaaayathwsntgg 975
         || ||.|| ||.|| || .||||.|| ||.|||||.|| ||.|| |||| ||||||.|| .||
Db    35 TGCATNAAGCAGCAAGGAGCCTNAATGTTAATCCCANGACCATNGGGGAAAAATGTCTCCAGG 94

Qy   976 acntgycarmgnwsnwsncarcaarwsnathaaarwsnytngcntggmgnccnmgnng---n 1032
         || ||.|| .||.||.|| || || .||.|| |||.||.||.|| ||.|| ||.||.||---n
Db    95 TCAYTGCAGAGACCTTCAGTCGGGG-CAGAGCCCCACCCACACAGAGTCCTGTGTGG--n 154

Qy  1033 aaatggttytgtyggnacnggnccnggnwsnytntcgtgtyggtncarccnmggayytngtn 1092
         || ||||.|||| ||||.||.||.||.||.||.|||.|| .|| |||| |||.||| |||||
Db   155 AAGTGGTTTGTTGTGGGCTGGGGCCCCAGGGCGTCTCTGTGCGTGTGCCAGCCTAGGGACTTGGTG 214

Qy  1093 ccntgyytnccng------------------------------tnaaywsngncngtn 1119
         || ||.||.||||------------------------------||||.||.||.||||
Db   215 CCCTGTGTTCCCAGCTGCTCCAGCCATGGCTGGGGAGGGTACCTGCTTCAGAGGGTACGG 274

Qy  1120 gcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggn 1179
         || ||.|| ||.|| ||.|| |||||.||||| ||.|||||.||.|| || ||.||.|||
Db   275 GCTTCAGAGGGTAGGAGGCCCCCAAGCAATTGGCAGCTTCCAGTGAGTTGAGCGTGTGGGT 334

Qy  1180 gcnaaraatwsnmgnathgargtntggagarccnccnaathmgnttycaraaraathtayggn 1239
         || || ||.||.||||.|| || || ||||| || ||.||.||.|| |||| ||.|||.|||
Db   335 GCAGAGATGTCAAGAATTGAGAATTGGAAGAACCCTGCCTAGATTTCAGAATATATATGGA 394

Qy  1240 aayccntggatggccnmgncaraartytyccngtnnggntgggnwsnwsntggmgnacnwsn 1299
         || || ||||||||||.||.|| |||| ||.|||.|||||||||.||.||.|||.||||.||
Db   395 AACACCTGGAAGTGTCCAGGCAAAAGTTTGCTGCAGGGGCCAGGAGCCCTCAGAGGGAACCCTCT 454

Qy  1300 gcnmgngtngtncaraargnaaygtnggntgggarccncncaymgngtnccnwsnggn 1359
         || ||.||.||.|| || |.||.||.||.|||||.||.||||.||.|| |||.||.|||
Db   455 CCTAGGGCCAGTGTGGAAGGGAAAATGTGGGG-CAGAGCCCCACACACAGAGTCCTGTGG 513

Qy  1360 gcnccnwsnwsnmgncngtnmgnmgnwsnccnccnws------------------n 1398
         || ||.||.||.||.||.||.||.||.||.|| ||.||------------------n
Db   514 GCACTGCCCTAGTGAGCTGTGGAGAGAGGCCACACNCCNWS------------------n 1398

Qy  1399 wsnmgnytncaraargnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacn 1458
         ||.||.||.|| || |.||.||.|| || .||.||.|| ||.|| ||||| || ||.||
Db   574 GGTAGACCCCCAGAATAGTAGATCAACAACCACCAACTTGCACTCTGTGCCTGGAAAAGCTACA 633

Qy  1459 gaycacncartgycarccngtnaargcngcngggnaatggarwsngtnccntayaaracngtn 1518
         .| || ||.|| ||.|| || ||.|| ||.||.||.|| || |||| || ||.|| ||.|||
Db   634 GACACTCAACATCAGCGTGAGAAAGCAGCCAGGAGGGAGGGCTGTACTCTGCAAAGCCACA 693

Qy  1519 gtngcngarytnacnacngtngngnaathtayytnaytcaytgycaygayytngaygtn 1578
         .||.|| || ||||.|| ||.||||| |||||.||.|| ||.|| ||.||.||.||.||
Db   694 GGGGCGGGAGCGTGCACAAGGACCATGGGAACCTATCCTGCATCACCATGACCTGAAATGTG 753

Qy  1579 mgncaygggngtnaarmgnaycayttyggngcnytnmgnttygaytgyccnacngnntty 1638
         .||.|| ||.||.|| ||.|| |||||.||.|| ||.|| ||.||.|||||.|| ||.||
Db   754 AGACATGGAGTCAAAGGAGATCATTATTGGAGCTTTAAAAATTTGACTGCCCACTGATTTT 813

Qy  1639 mgnacntayatgggnccngt-ncnnyntgyttyggncaartyttyccnttyggnacngc 1697
         .|| || |||||||.||.||-||.||.||.|||||||||| .|||||.|||||||| ||
Db   814 CAGACTTGCCATGGACCCCTGTAACCCCTTTGTTTTTGGGCAATTTCTCCCTTTGGAACAGC 873

Qy  1698 ngtnttyacncnaartgyytntayy 1720
         ||.|||||.||.|| ||.|||||||
Db   874 TGTATTTACCCAAATACCTGTACC 896
```

RESULT 8
LOCUS   AG086951   717 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-085H08.F, genomic survey sequence
AG086951

ACCESSION   AG086951
VERSION   AG086951.1 GI:16638753
KEYWORDS   GSS; GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
             BAC Library clone:PTB-085H08.F.
ORGANISM   Pan troglodytes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   BAC end sequences of Library PTB
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 717)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL   Direct Submission (02-AUG-2001) Asao Fujiyama, The Institute of Physical
             and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
             1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jP/,
             Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
             was generated during the R&D process and may have higher chance of
             clone tracking errors.
FEATURES   Location/Qualifiers
     source   1..717
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-085H08.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
     PRIMERS
                     Sequencing: -21M13
     LIBRARY
                     Vector   : pKS145
                     R.Site 1 : SacI
                     R.Site 2 : SacI
BASE COUNT   175 a   189 c   195 g   158 t
ORIGIN

Query Match 18.1%; Score 313; DB 12; Length 717;
Best Local Similarity 49.1%; Pred. No. 2.8e-61;
Matches 299; Conservative 113; Mismatches 195; Indels 2; Gaps 1;

Qy   1098 ygtnccngtnaaywsngcngtngcnwsngargggngcnwsnccnaaarccntggcaarytncc 1157
            ::|| ::|| ||: |:|| :| ||||::: ::|| :||:|||| :|| ||||| ::|||| |
Db   89 CGTCCGCGGAGCTCAGGCNGTGGGCTTCAGAGAGGGTGCAAGCCCTNAGCCTTGGCAGCTTCC 148

Qy   1158 nwsngggntngarccngtngggncnaataarwsnngnathgargtntggaarccncncnat 1217
            :|| ||:|| :|| ||:: :|| ||||:||| || ||:|| |||| : :|||| |:|||
Db   149 ACATGGTGTGTAAGCCTNGTGAGTGCACAGAAANTCAAGAACTGGGTTTGGGAACCTCTGCC 208

Qy   1218 hmgnttycaararathtayggnaayccntggatgcccnmgncaraarttygcngtnggngt 1277
            ::||||:: |: ||:|:|||:| |||::|| :|| || ||||:||:||:||:||:||||:
Db   209 TAGATTTCAGAGGAATATATGGTAANTAACCTGGATGCCCAGACACAGAATTTTGCTGTAGGGA 268

Qy   1278 nggnwsnwsntggmgnacnwsngcnmngtngtncaararggnaaygtnggntgggarcc 1337
            :::|| :||| : : |||| :| :|| ||::|| ||| |::| ||:||| ||||||::
Db   269 GGGGGCCTCATGAAAAACCTCTGCTAGGGACGAGTGTGGGAGAGAAATGTGGGGTCAGAGACC 328

Qy   1338 ncncncaymgntgcncnwsnggnggncncnwsnsnmgngncngtnmgngmgnwswsncnws 1397
            : || :|| |||| || :||::||: :|:: ||:| |: ||||| : ||| :|||:|||
Db   329 ATCACACAGAGCTCCTACYTGGCGCACCGCCTAGTGGACGTGTGAAAAGTGGGCACTAYT 388

Qy   1398 nwsnmgnytncaararggnmgnwsnacngaywsnytncarcaygtnccngaarararwsnac 1457
            :|| :|| || |: ||::| |||| :||:|| :||||: |: ||||| :||||| :||:
Db   389 CTCCAGACCCAGAAAYGGTAGAYCCACTGACACAGCTTGCGCACATGCACCTGGAAAAAGCCAC 448

Qy   1458 ngayacncaartgycaycnccngtnaargcngcnggnatggarwsngtnccntayaaracngt 1517
            :|:|| :||:||: :|:: ||::||:: |||: :||||: | ||:||::||: :|||::|||
Db   458 TGGTCTGGACCACCAAGTCCCCATGTGTYGTGTCAGCCTAGGACGTTGCTGCCTGTGTTCCCA 1072

Db   449 AGACACTTCAATGCCAGCCTGTGAAAGCAGCCTGGGAGGAGGGAGGAGGACTGTTGACCATGCAAGACCAC 508

Qy   1518 ngtngncngaryytnacnaaaracngtnggnnathtayytnytncayytgycayygayytngayaygt 1577
            :|| ::|||| ||:::|||:| | || ::|| ::|::||:|:::|| || :||:||| :|||| |:|
Db   509 AGGGGCAGAGCTGCCCAAGACACCATGGGAACCCACCTCTTGCATCACTCTTGCATCATCATGACCTGGAATGT 568

Qy   1578 nmgncaygngtnaarmgnaaycaytyggngcnytngnctygaytgaycgcnacngggntt 1637
            :|| :|:|| || | |: ||: ::||:|| |: | |: ::|::|||:|:| :|||:|||
Db   569 GAGACATGGAGGACCAAAAGGATCATTTTGGGAGGCTTTAAGGATTTGCACTTCCCTGCTCTGCCTGGGATT 628

Qy   1638 ymgnacntayatgggncngt--ncnytntgytyggncaartyttyccnttyggnacn 1695
            :|| :| ::|| :||| || ::|::| |:| ||| ||:|||: |: || ||| |:|
Db   629 TCAGACTTGCATGGGCCCTGTAGCCCTGTCTTTTTTTTTTTTGGCCAATTTCCCCATTTGGAATG 688

Qy   1696 gcngtntty 1704
            |:|: ::|
Db   689 GGGTTAATT 697

RESULT   9
BI754555/c
LOCUS   BI754555   736 bp   mRNA   linear   EST 25-SEP-2001
DEFINITION   603023384F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194028 5',
             mRNA sequence.
ACCESSION   BI754555
VERSION   BI754555.1 GI:15746133
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 736)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11485 row: g column: 21
             High quality sequence stop: 733.
FEATURES   Location/Qualifiers
     source   1..736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5194028"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned [EcoRV site is destroyed
                     upon cloning]. Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT   163 a   196 c   185 g   192 t
ORIGIN

Query Match 17.7%; Score 306.6; DB 10; Length 736;
Best Local Similarity 44.1%; Pred. No. 8.8e-60;
Matches 306; Conservative 122; Mismatches 266; Indels 0; Gaps 0;

Qy   1013 tngcntggmgnccnmgnmgnaartggttytgtygggnacngngnwsnytntgtgyg 1072
            ::|| | || ||::|||:| ::|||:|||| |::|| |||::|| |: ||::|
Db   635 TGGTCTGGACCACCAAGTCCCCATGTGTYGTGTCAGCCTAGGACGTTGCTGCCTGTGTTCCCA 694

COMMENT   was generated during the R&D process and may have higher chance of
           clone tracking errors.
PRIMERS   Sequencing: M13Rev
LIBRARY   Vector   : pKS145
           R.Site 1   : SacI
           R.Site 2   : SacI
FEATURES   Location/Qualifiers
     source   1..654
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-119B01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT   165 a   174 c   180 g   135 t
ORIGIN

```
Qy   1073 tncarccnmngayytngtncccntgtygttngccngtgtngcnwsngaargngg 1132
Db    634 GCCACTCCAGCCCATTGCTAAAAGGCGCAAGTAGCACCTAGCCACCCATGGTTTCAGAGAGT 575
Qy   1133 cnwsnccnaarccntggcaarytnccnwsngggntngarccngtngngcnaaaraarwsnm 1192
Db    574 GCAAACCCAAACCTTGGCAGCTNACCANTGTGGTTGTTGAGCCGTGTTACAGAGTCAA 515
Qy   1193 gnathgargtntgggarccnccnathmgnttycaraaarathtayygnaaycntggatgc 1252
Db    514 GAATTGAGGTTGGGAACCTCCATCTAGATTTCAGAAGATGATNGAAAAATGCCTGGATGC 455
Qy   1253 cnmgncaraarttygcngtngggntnggarccnccncncncaymgngtnccnwsnsntggmgnacnwsngcnmgngtngtnc 1312
Db    454 CCAGGCAAAAGTTTGCTCACGGGAGGGGGCCCCTCTTGGAGAACCTCTGCTAGGGCAGTGC 395
Qy   1313 araarggnaaygtngngntnccnwsngngncnccnwsnsnm 1372
Db    394 AGAAGGGAAATGTGGGGTCAGAGCCCAACACAGTATNCCCTTACTCGGGGGTACTGCCTAGTA 335
Qy   1373 gngcngtnmgngnmgnwsncncncnwsnwsnmgnytncaraargwnmgnsnacngaywsny 1432
Db    334 GAGCTGTAAGAGNAAGGCCACCAATCCTCCAAACATCAGAANTGGTTAGATACATTGACAGCT 275
Qy   1433 tncarcaytncngararaarwsnacngayacncaartgycarccngtnaargcngcngna 1492
Db    274 TGTACAAGGTGCCTGGAAAACCCGAGACACTCAATGCCAGCCATGAAGAGAGCCAGCA 215
Qy   1493 tggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayy 1552
Db    214 GGGGGAGCTATGCCCTGCCAAAGCACAGGGGTGGAGCGTGCCTACGACTATGGGAACACACC 155
Qy   1553 tnyttncaytcaytgcaygayytngaygtnmgncaysggntnaarmgngaycaytyggngncny 1612
Db    154 TCTTCGGATCACTGGATTGGGATNATGAGACANTGCAGTCAACATTGAGAAATTCACTTGGGCTT 95
Qy   1613 tnmgnttygaytgyccnacncngnttymgnacntayatgggnccngtncnytntgyttyg 1672
Db     94 CAAAAATTTGACTGCCCCGCTGGATTTTGGACTTGCAATGGGTTCCTGTAACCCTTTGTTTT 35
Qy   1673 gncarttyycnttyggnacncngtnttyac 1706
Db     34 GGCCAATTGTCCCCATTTGGANTGGCCGTATTCAC 1
```

RESULT 10
LOCUS   AG112666   654 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-119B01.R, genomic survey sequence.
ACCESSION   AG112666
VERSION   AG112666.1 GI:16733185
KEYWORDS   GSS; GSS (genome survey sequence).
SOURCE   Pan troglodytes
            Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-119B01.R.
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   BAC end sequences of library PTB
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 654)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   clones are derived from the chimpanzee BAC library PTB This BAC end

Query Match Best Local Similarity 17.4%; Score 300.4; DB 12; Length 654;
Matches 296; Conservative 49.3%; Pred. No. 2.2e-58; Mismatches 191; Indels 2; Gaps 2;

```
Qy   1114 gcngtngcnwsngargngngcnwsnccnaarccntggccarytnccnwsngggntngarccn 1173
Db     50 GCTGTAGCTTCACGAGGTGCCAAGCCTCAAGCCCCTTGGCAGCTTCACATGATTGTTAAGCCT 109
Qy   1174 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraaarath 1233
Db    110 GTGAATGCACAGAGAGTCAAGAAATTGGGGTTTGGGAACCTCCACTTAGATTTCAGAAAATG 169
Qy   1234 tayggnaayccntggtatgccnmgncaraarttygcngtngggnwsnwsntggmgn 1293
Db    170 TATGGAAACTCCTGGATCCCAGGCAGAGAGTTTGCTCCTGGGGTGGGGTCCTCATGGAGA 229
Qy   1294 acnwsncngmgntngtncarcaytgaygaaraargngnaaygtngggngttgacagacna 1353
Db    230 AACTCTGCTAGGGCACTGCAGAAAGAAAAAACGTGGGGGTGGGGGGGGGTTCGGGGGCCCCA 289
Qy   1354 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsncncncnwsnwsnmgnytncaraar 1413
Db    290 ACTGGGGCACTGCCTAGG-GAGCCGTGAGAAGAGGGCCACTGTCCTCCAGACCCCAGAAT 348
Qy   1414 ggnmgnwsnacngaywsnytncarcaygtnccngararaarwsnacngayacncaartgycar 1473
Db    349 CGTAGATCCACTGACAGCTTGCCCATGTACCTGGAAAAGCCAGACACTCAACACCAC 408
Qy   1474 ccngtnaargcngcngcngnatggarwsngtnccntayaaracngtngtngcnga 1533
Db    409 CCCATGAAGGCAACTGGGAGAGAGCCGTCACCTGCAAAGCCACGGGGTGAAACTGCCC 468
Qy   1534 aaracngtnggnathtcaytgcaygayytngaygtnmgncaysggngtnaar 1593
Db    469 AAGACTAGGGGGAAACCTACCTGTTCACATCAGCANTGGTTAGATACACCTGG 528
Qy   1594 mgngaycaytyggngcnytnmgnttygaytgyccnacncngnttymgnacntayatgggn 1653
Db    529 GGAGATCACTTTGGAGCTTTAAGATTTGACTGCCCCACGGGATTTCAAACTTGCCATGGGC 588
Qy   1654 ccngt-nccnytntcgyttyggncarttytytcyccntttyggnacncngtnttyacncartg 1712
Db    589 CCTGTCAGCCCCTTAGTTATGGCCAATCTCTCCCCATTTGGAAATGGCTGCACTTCACCCATG 648
Qy   1713 y 1713
Db    649 T 649
```

RESULT 11
LOCUS   AG064424
DEFINITION Pan troglodytes DNA, clone: PTB-053H24.R,   676 bp   DNA   linear   GSS 03-NOV-2001
            genomic survey sequence

ACCESSION   AG064424
VERSION   AG064424.1   GI:16616226
KEYWORDS   GSS; GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
             BAC Library clone:PTB-053H24.R.
ORGANISM   Pan troglodytes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Unpublished
JOURNAL
REFERENCE   2 (bases 1 to 676)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
             Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
             and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
             1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
             Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   BAC end sequences of Library PTB
             Clones are derived from the chimpanzee BAC library PTB This BAC end
             was generated during the R&D process and may have higher chance of
             clone tracking errors.
FEATURES   Location/Qualifiers
     source   1..676
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /sex="male"
                     /clone="PTB-053H24.R"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
PRIMERS
             Sequencing: M13Rev
LIBRARY
             Vector   : pKS145
             R.Site 1   : SacI
             R.Site 2   : SacI
BASE COUNT   164 a   165 c   191 g   156 t
ORIGIN

Query Match   17.1%; Score 296; DB 12; Length 676;
Best Local Similarity   47.9%; Pred. No. 2.3e-57;
Matches 294; Conservative 117; Mismatches 201; Indels 2; Gaps 2;

Qy   1114 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
            | ||:| | ||:| :|||||| :||||| | |:| | |:|| ||| ||: | |||:|
Db   63 GCCATGGCTTCTAAAGGTGGCAAGCCTCCAGCCTTCCACATGAGTTGAGCTG 122

Qy   1174 gtnggngcnaaraarwsnmngnathgargtntggaarccnccnacnathmgnttycaraarath 1233
            |:|| | ||||| |:||| ||| || : | |:||||| ||| || || :| :||
Db   123 GCAAGTGCACAGAAGTCAAGAACTGGGGTTTAGGAACCTCTACCTAGATTTCAGAGATG 182

Qy   1234 tayggnaayccntggatgccnmgncaraarttygcngtnggtnggnwsnsntggngmgn 1293
            |||:|| :||| ||||||| | ||||||| ||| ||||| |||:| :|||:|| |
Db   183 TATGGAAATGCCTGGATGTCCAGGCGAGAGAGTTTCCGCCAGGACGGGGTCCTTGGGAGA 242

Qy   1294 acnwsngcnmgngtngtncaraaraggnaayngtnggntnggngaycncnccaymgngtncccn 1353
            | |||||| ||||| | ||||:||| |||||| | ||||| ||||:|||| ||:||
Db   243 ACCTCTGCTAGGGACGTGGAAGGGAGAAATGTGGGTTCCAGCCCAGAGTCCCT 302

Qy   1354 wsnggngcnccnwsnsmngncngtnmngnngnwsnccnwsnmgnytncaaratccn 1413
            :|:||| ||| :||| ||||| ||| ||||||| || || :| ||| || || | |
Db   303 ACTGGGGCCACCATCTAGTGGAGCTGGAGAAG-AGGGCACTGACCTCACCAGGAAT 361

Qy   1414 ggnmgnwsnacngayvsnytncacracaygtnccngaraarwsnacngayacncartygcar 1473
            | ||:|| | | ||| | | ||| || |:|| |:|| |: | || ||:| | |
Db   362 GGTAGATCCACTGACACGTTCACCATTGTCCCTGAAAAAAGCTGTAAACACTCAACACAG 421

Qy   1474 ccngtnaargcngcngnaatggaarwsngtncncngaarytnacn 1533
            || || || |||||| || |:|| ||| | ||| ||||| |
Db   422 CCCATGAAAGCATCTGGGAGGAGGAGGAGGTTGTTACCCTGCAGATCCACAGGGGCGCGGGAGCTGCCC 481

RESULT 12
LOCUS   AG062447   691 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-050M20.F, genomic survey sequence.
ACCESSION   AG062447
VERSION   AG062447.1   GI:16614249
KEYWORDS   GSS; GSS (genome survey sequence).
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-050M20.F.
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (sites)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Unpublished
JOURNAL
REFERENCE   2 (bases 1 to 691)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   BAC end sequences of Library PTB
            Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
FEATURES   Location/Qualifiers
     source   1..691
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /sex="male"
                     /clone="PTB-050M20.F"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
PRIMERS
            Sequencing: -21M13
LIBRARY
            Vector   : pKS145
            R.Site 1   : SacI
            R.Site 2   : SacI
BASE COUNT   160 a   167 c   201 g   162 t   1 others
ORIGIN

Query Match   16.6%; Score 287.4; DB 12; Length 691;
Best Local Similarity   49.0%; Pred. No. 2.3e-55;
Matches 196; Conservative 110; Mismatches 196; Indels 7; Gaps 3;

Qy   1534 aaracngtnggnathtayytnytncaytgtycayvgayytngaygtnmgncntayatgggn 1593
            ||:||| | | |||||||| |||| || |:|| | | |||| | | :|| |||||||||
Db   482 AAGACCATGGGAGCCCACCTCTTGCATCACAGTGTGACCTGGATGTGAGACATGGAGTCAAA 541

Qy   1594 mgngaycaytgyggngcnytnmgnttygaytgyccnacnggngtctymgnaacntayatgggg 1653
            |||| || |:|||||| | ||| | | || | || | | ||| ||| | |||||||
Db   542 GGAGATCAtttgtaGTTTTAAGAttTGACTGCCTCCGCCTGAttttggACTAGCATGGGG 601

Qy   1654 ccngt-nccnytntgytyggncarttytyccnttyggnacngcngtnttyacncartg 1712
            |||| | ||| |||||||| ||| | | || | | || | || || ||| || ||
Db   602 CCTGTAGCCCCTTGTTTGGCCAATTCTCCCAATTCGAATGGCTGTTATTTACCCAATG 661

Qy   1713 yytntayytncayt 1726
            :|| ||: |||
Db   662 CCTGTACCCCATT 675

Qy   1158 nwsnggngtngatccngtngtgngngcnaaaraarwsnmgnathgargtntgggarccnccnat 1217
           ::|: :|| ||::| ||:| :|||:|| ||| || |||: |: | | ||:| || |||
Db   142 ATGTGGTGTTGAGCCNGTGACTGCACAGAAGA-----ATCGGGTGGTTTGGGAACCTCCACT 196

Qy   1218 hmgnttycaraarathtayggnaayccntggatgccnmgcnacataaartyygcngtngaggngt 1277
           ||:| ||||| ||||| || | ::| || || || | ||| |||| |||| ||
Db   197 TAGATTTCAGAAGACATGTAAGAAATACCTGGATGCCATGCCAAAANGTTTGCTGTAGGGTC 256

Qy   1278 nggnwswsntggngnacnwsngncnmgngttngtncaaraarggnaaaygtghggntgggarcc 1337
           : :| || ||| || | || : | :|| | | | | :| | ||
Db   257 AGGGCCCTCATGGAGACNTTCTGCTAGGGCGTGTGAGAAGGGAAAATGTGAGGTCAGAGGC 316

Qy   1338 ncncaaymgngtnccnwswsnmgncngtnmgmgnwsnccnccnws 1397
           :| | | | | ||::|:| :|| : | :||| |||| :
Db   317 CCCACACAGAGTCCCTACTGGGGCNACCCACCTGAGCTGTGAGACCTGAGAGAGGC 376

Qy   1398 nwsnmgnytncararggnaaatgrgwsnac 1457
           :|: |||:||:|||:| ||:| :|| |:|
Db   377 CTCCAGACTCCAGAATNGGTAGANCCACTGAGATCCNCATGTGCCTGAAAAGCCAC 436

Qy   1458 ngayacncartgycarccngtnaarcgncgngnnatggaarwsngtnccntayaaracngt 1517
           :|| ||| || | || | || :| | || | ||| | :||| |:| |||
Db   437 AGACACTCAATGCCACCCATGAACACGGCGGGGAAGAGCCTGTGACCTACAAGGCCAC 496

Qy   1518 ngtngcngaryt-nacnaaracngtngngnaathtayytnytncaytcgycaygayytngayg 1576
           ||| || |::| || || |||:| | | :|:| :|| | | ||| :| |||| |
Db   497 AGGGGCAGAGCTGTTCAAGACAATGGGAACCCACCCCCGTTGCCTCAGTGTGACCTGGTTG 556

Qy   1577 tnmgncaygggngtnaarmgngaycaytcgygggncgnytnmgngnttygaytgyccnacanggnt 1636
           :|| | || | || | | | | ||:|| ||| :| ||| | | || ||
Db   557 TGAGATATGGAGCTCAGAGGATCAATTGTAGAGCTTTAAGATTTGACTGTGTGGACTTGG 616

Qy   1637 tymgnacntayatgggngnccngt-nccnytntgyttyggncartttytyccnttygngnacn 1695
           ||: | | ||: | || || || ||| ||| :|| | :||::| | |:|| ||
Db   617 TTGGACTTGCATGGGCGTCCCTTGTTTGGCCAATTTCTCCCATTTGTAATG 676

Qy   1696 gcngtnttyacnca 1709
           ||| ||||| ||
Db   677 GCTGTATTTACCA 690

Qy   1114 gcngtngcnwsngargngngncnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
           :| || || || |||| |:| | | | : || :|| :|| ||| || ||| |
Db   110 GCCATGGCTTCAGATGGTGCAAGCACCAAGTCTGTGGTGGCTTCCACATGGCCATTGAGCCT 169

Qy   1174 gtnggngcnaaaraarwsnmgnnathgargtntgggarccnccncnathmgnttycaraarath 1233
           ||:|| || | | | || | || | ||| ::| || || | |:| ||:||| |||
Db   170 GCGGGTGCACAGAAGAGTCAAGAAAATTGACGTTTGGAAACCTCCACTGAATTTCAGAACATG 229

Qy   1234 tayggnaayccntggatgccnmgncararartycyngtngnttggggnnwsnwsntggmgn 1293
           ||:|| || | || ||| |: | || | :|:|| | | ||| ::| |||
Db   230 TATGGAAATGCCTGGATGCCCAGGCGAGCAGAAGTTTCGTCCACCTGAATTTCAGAACATG 289

Qy   1294 acnwsngcnmgngtngtncararggnaaygtnggnttgggnwsnwsntggmgn 1353
           || :| ||: | | ||| |: | ||::| ||| |: | |||
Db   290 ACCTCTGCTNAGGGCAGTGCACAGAAGAGAAAATGTGGGGTCGGAGCCCCACACAGGTGCCT 349

Qy   1354 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccccncwsnwsnmgnytncaraar 1413
           :| :| || || || |||| || | || ||| | : |:|| ||
Db   350 ACTNGGG-ATCACCTTAGTGGCNCTTAGTGGAAGACAGCACCACTGTTCTTCCAGACCCAGAAT 408

Qy   1414 ggnmgnwsnacngaywsnytncarcaygctgytgyccnacgayacancartgycar 1473
           | : |||:: |: : || : | | | || || || :| | |
Db   409 GCTAGATCCACTGACACTGGC--ACCTGTGCCTGGAAAAGCTGCAGGTACTCAACATCAG 466

Qy   1474 ccngtnaaracgncgngnnatggaarwsngtnccncanggnttyngncancartgycacncar 1533
           | || || | | | | || || ||:|| | ::| || : || | | ||
Db   467 CCCATGAAAAGCAACTCGGGAGGGAGGCCTGTACCCTGCAAAGCAACAGAGGTGCAGTTCCC 526

Qy   1534 aaracngtngnnathtayytnytncaytcgytcaygaygayytngayg 1593
           || | || | |:| || || :| | :| || ||::||
Db   527 GAGACCAATGGGAACCTACCTCTTGCATCAGGTGGAGTGCGAGTGAGACAGGATTCAAA 586

Qy   1594 mgngaycaytcyggngcnytnmgnttygaytgyccnacngnttymgnacntayatgggn 1653
           || | || |:|:| | :| |||: || || || ||:| || | | ||:|||
Db   587 AGAGATACATTATTGGAGCTTTAAGATTTGACAGCACACTGATTTTGGACTTGCATGGGC 646

Qy   1654 ccngt-nccnytntgyttyggncartttytyccntcyggnacncgntnttyacncaytg 1712
           :| || ||| ||| :|| | :||::| || |:|| || | ||| |: | |
Db   647 CCTGTAGCCCTTTTGTTTGTTTGGCCAATTTCTCCCATTTGGAATGGTTGCATTTACCTAATG 706

Qy   1713 yytntayytncayt 1726
           :|: |::||||
Db   707 TCTGTATTCCCATT 720

LIBRARY
    Vector   : pKS145
    R.Site 1   : SacI
    R.Site 2   : SacI.
FEATURES   Location/Qualifiers
    source   1..720
                      /organism="Pan troglodytes"
                      /db_xref="taxon:9598"
                      /clone="PTB-124K01.R"
                      /sex="male"
                      /cell_type="lymphoblast"
                      /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT   181 a   186 c   187 g   165 t   1 others
ORIGIN

Query Match   16.5%; Score 286.2; DB 12; Length 720;
Best Local Similarity 48.5%; Pred. No. 4.5e-55;
Matches 298; Conservative 112; Mismatches 200; Indels 4; Gaps 3;

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
  COMMENT   Other_GSSs: CIT-HSP-2337G10.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2337G10"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT        160 a      146 c      172 g      145 t
ORIGIN

Query Match          16.4%;   Score 283.8;   DB 12;   Length 623;
Best Local Similarity  48.1%;   Pred. No. 1.5e-54;
Matches  286; Conservative  111; Mismatches  196; Indels  2; Gaps  2;

Qy  1138 ccnaarccntggccarytnccnwsngggntngagarccngtnggngcnaaaraarwsnmgnath 1197
         | ||||||| ::| ||||||| || |||::|||||||| |:||||| |||:||||||||||||||
Db    26 CTCAAGCTTTGGCAGCTTCCACATnGGTGTTGAGTCTGCGGGTACACCAAAGTCAAGAAACT 85

Qy  1198 gargtntg-ggarccncconathmgnttycaararathtayggnaayccntggatgccnmg 1256
         ||||||||  ||  ||:||||||||||||||||||:|||||||||||||:||||||||:||
Db    86 GAGGTTTGAGGAACCTCCACCTACATTTCAGAGAGATGTATGGAAAATGCCTGGARGTCCAG 145

Qy  1257 ncaraarttygcngtnggngtnggngwsnwsntggmgnacnwsngcnmgngtngtncaraa 1316
         |:||| |||| ||||:|| |||||||| ||||| | ||| | | ||| | ||||||||||
Db   146 GCAAAAGTTTTGCTTCCAGGGGCACAGGGCCCAGGATCGAGAACCTCTGCTAGGCAGTGCGAA 205

Qy  1317 rggnaaygtnggngntggaarccncnccnccaymgngcnccnwsngggcncconwsnwsnmgngc 1376
         :|| || :|||||||||||||| |||||||| |:||||| | ||| :|| |||||||||||
Db   206 GGGAAANTGTGGGGTTGGAACCCGTCACAGAGAGTCTCTACGGGGCACTCCCTAGTGGGAGC 265

Qy  1377 ngtnmgnmgnwsnccnccnwsnwsnmgnytncaraargngnmgnwsnacngaywsnytnca 1436
         |:||||||  ||| :||:| ||| |:|  |||:|||||||||:||||||||| ||:||
Db   266 TGTGCAGAAGAGGGCCCATTGTCCTCCGACCCTAGAATGGTAGATCTGCGACAGCTTGCA 325

Qy  1437 rcaygtnccngararwsnacngayacncaartgycarccncngtnaargcngcnggnatgga 1496
         ||||| |||:|||||||| ||||||||||||| ||| ||||||:|| |:||||| | |||
Db   326 TTAnGCAACTGGGAAAAGCCACACAGACACTCAACACACCCATGAAAACAGCCAGGAGGG 385

Qy  1497 rwsngtnccntayaaracngtngcngarytnacnaaaracngtngnathtayytnyt 1556
         |:| | |:||| | |:| | ||| | || |||||||::|| || |||| ::|||:|
Db   386 AGCCATACCCTGCAAAGCCACAGGGCCAGAGTGGCCTGAGCAAACCACCACTCCT 445

Qy  1557 ncaytgycaygayytngaytgmgncayggngtnaarngaycayttygggngcnytnmg 1616
         |:| ||:|| :|| ||| | |||| | |||| ||||| || | || ||||| ||:
Db   446 GCATCAGGCTGACCCAGATGTGACACATGGTGTCAAAAGAGATCATTTTGGGACTTTAAG 505
```

```
Qy  1617 nttygaytgycnacnggnttymgnacntayatgggnccngt-nccnytntgyttyggnc 1675
         ||:||:||  | |  |||:||| :| | ||| || || || | |: | :||::||||||| |
Db   506 ATTTGACTGCCCTGCTGGATTTCAGAACTTGTAAAGGGGCCAGTAGCCCTTTCTTTTGGCC 565

Qy  1676 arttyttcycnttyggnacngngtnttyacncartgyyntatyytncaytgyat 1730
         || ||| ::| ||| |  ::||||| |||||||||| ||||||| |||:|||||| || ||
Db   566 AATTTCTCCCATTTGGAATGGCTGTGATTTTACCCAATGCCCTGTATTCCCATTGTAT 620
```

```
RESULT 15
AG097258
LOCUS       AG097258              677 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-098001.R, genomic survey sequence.
ACCESSION   AG097258
VERSION     AG097258.1  GI:16717775
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-098001.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 677)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
  COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: M13Rev
            LIBRARY
              Vector     : pKS145
              R.Site 1   : SacI
              R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..677
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-098001.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT        178 a      175 c      176 g      147 t      1 others
ORIGIN

Query Match          16.4%;   Score 283.2;   DB 12;   Length 677;
Best Local Similarity  48.8%;   Pred. No. 2.1e-54;
Matches  294; Conservative  113; Mismatches  187; Indels  9; Gaps  3;

Qy  1114 gcngtngcnwsngartggngngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
         || ||: | ::||  | |||||||| ||:| |||:|| :|||||| || | ||:||||:|||
Db    68 GCTGTGGGCTTCAGAAGGTAGAAGCCCCAACCCTTGGCAGCTTCCACATGTGTGTTGAGCCT 127

Qy  1174 gtnggngcnaaaraarwsnmgnathgargtntgggarccncnccnaathmgnttycaararath 1233
         ||:|||||||||||||||| |||| ||||||:|||| |||||||||||||:||||||||||:
Db   128 ACAGGTGCACAGAAGTCACAAATTGAGGTTTGTGAACCTCCACCTAGATTT-AGACGATG 186

Qy  1234 tayggnaaayccntggatgccnmgncaraarttygcngtnggngtnggngwsnwsntggmgn 1293
         || || |||:||||||| | || | |:||| ||||||||:|| ||||||  ||||| |||:|
Db   187 TATGGAAAACACCTGGATGCCCGGTAAAAGTTTGCTGCGGGGTGGAGGCCATCATGGAGA 246
```

```
Qy  1138 ccnaarccntggccarytnccnwsngggcnaaaraarwsnmgnath 1197
         | ||||||| ::| |||||||| ||
Db  1197

Best Local Similarity  48.1%;
Matches  286; Conservative  111;
```

```
Qy   1294 acnwsngcnmgntngtncaraarggnaaygtnggntggarccncncaymgngtnccn 1353
          || |:: |  ||| | :|:| ||||:|| || |||||||  |:||||||| ||
Db    247 ACCTCTTCTTAGGGCAGTACGGAAAGGAAAATGCAGGGTTGGAGCCCCAAACAGAGTCCCT 306

Qy   1354 wsnggngcnccnwsnwsnmngngcngtnmgngnwsncccnccnwsnwsnmgnytncaraar 1413
          :  ::|  | |  ||:: | ||| ||||||| ||:|||| | |: |||||
Db    307 AATGGCCGCACTGCCTAGTGAAGCTGTAAGAAGAGGACCACCATCCTCCAGACCCCAGAAT 366

Qy   1414 ggnmgnwsnacngaywsnaytncarcaygtnccngaraarwsnacngayacncaartgycar 1473
          || |:|  | | ||::  :| |: :|||:  | ||||||:||| |||:||||||:|::|
Db    367 GGTAGATCCAACCTACAGCTTGCAACATCGCACCTGGAAAAGCTGCAGACACTCAGCGCCAG 426

Qy   1474 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1533
          ||  ||||:| |||||||||||| ||||| ||:|||:| |||  ||| |||:| |||:|
Db    427 CCCATGAAAGCAGCCAGGAGGGACTGTCTCCTGCAAATCCACAGGGGCAGAGCTGCCC 486

Qy   1534 aaracngtnggnathtayytnnytncaytgycaygayytngaygtnmgncaygggngtnaar 1593
          |||  | ||  || | ||:|| ||: || | :|  |:|| ||:|| | ||| | |||
Db    487 AAGACCAATGGGAACCCACCTCTTCCATCAAACGTGACCTGGATGTGAGGCATGGAGTCAAA 546

Qy   1594 mgngaycayttygngcnytnmgnttygaytgyccnacngngnttymgnacntayatgggn 1653
          |  :|| ||  | |: |    :|| :||| ::||::| |: |   |||||: ||:|||
Db    547 GGAGATCATNTTGGAGC-------TTTGATTGCCCTGCTGGGATTTTGGACTTGCATGGGC 599

Qy   1654 ccngt-nccnytntgyttyggncarttytyccnttyggnacngcngtnttyacncaartg 1712
          || | || :| || | ||| :|| ::|||:| :|| | ||: |||:|||| |||::||
Db    600 CCTGTATCCCTGTGATTTGAGCAATTTATCCCATTGAAATGGCTGTATTTACCCAATG 659

Qy   1713 yyt 1715
          ::|
Db    660 CCT 662
```